

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

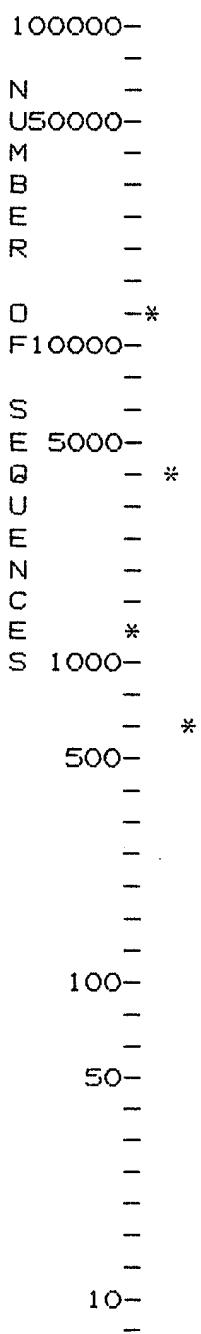
As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

LOW
07/344258

Results file 1ow344-fig1-pir.res made by maryh on Wed 17 Apr 91 11:19:26-PDT.

Query sequence being compared: L0W344-FIG1.PEP
Number of sequences searched: 17731
Number of scores above cutoff: 3826

Results of the initial comparison of LOW344-FIG1.PEP with:
Data bank : PIR 25.0, all entries



-CTGGCCTT-GCTCGAGGC- --GATCAGCTGGCGCTGTGCCACAGCC----CGGCCTTGCGCCAGG-GTTGG
 550 560 570 580 590 600
 610 620 630 640 650 660
 ----TGTAAAGGCAGGCCGCCGGGCCAGCTTGCGCACCAGGTGTTCCGGTAACCACTCACACGGCAGCAAGT
 610 620 630 640 650 660 670
 GCGCGGATCAAGGCG--CGGTGCGGCCAGGCTGGCGCGTGCGCCCTATTA-TCCCAGCGTCGC-GTT
 610 620 630 640 650 660 670
 670 680 690 700 710 720 730
 CAGCGC--GATGGTGAGCGAGGCCAG-GCCGC-CATTTTGAGTCCGAAGCTTGAGGCCCTCACGGGTTG-TA
 680 690 700 710 720 730 740
 GAGCGCCGGCGTTGGCGCAGCGGGCAGCACGGGGTTGGG--CGAGGACGGGGTCGCAACAATGTGATG
 740 750 760 770 780 790
 TTGGTCAC---AGCGATGATACTG-ACGAATTGAGCTATC--TC-ACCGCCCTGCTGCGCGG-ATA-CCTCA
 750 760 770 780 790 800 810
 GCGGTAACGCTGGCCTGGCGCCCTGTTGACTCGGGCGCCCGTTGGCCGCCCTGC-GGGCGGCACAGGCGCA
 800 810 820 830 840 850 860
 800-GGTCTAGACCACATCCCGCACAGTGCAGTTGGT-CTAGAAGATAATGCGAGTGCATCACCGCTCCTGGG
 820 830 840 850 860 870
 ACTGGACGAGGGCGC-GCAGGCCTATGCG--CGGTGCTGCAGGACAA-GC-TG-GCCGAAGTGGTC---GG
 870 880 890 900 910 920 930
 CATCCGTT-CGTGGCAAACACGGGCTCTTGTATCAAGGGCGCT-CATCGACCAAGGGCTACATGAAAC--AAA
 900 910 920 930
 CGCCATTACGAGGCGGCCAC-GCGC---AGGCGCTGCAT--ACTGCGG-TGGAAGACACGGAGA
 880 890 900 910 920 930
 940 950 960 970 980 990 1000
 TC-CTCG--TTTC-GAATGACTGCGCTTGG-GGTTTCGAGCTATG--TCACCAACATCATGGACGTGATG
 940 950 960 970 980 990 1000
 TCGCCCGGCGTTGGCCAGCATCGCGCGCGCGCGCGC-AGGCCTGGACAGGCCACGGCGATGTGCTG
 1010 1020 1030 1040 1050 1060 1070
 GAT-CGCGTGAACCCC--GA-CGGGATGGCCT---TCATTC-CACTGAGAGTGATCCCATTCTACGAGAGAA
 1010 1020 1030 1040 1050 1060 1070
 CATGCGCAGGCTGCCCTGGAGCGCGCGCGCAGGCGCAGGCGAGGCAGGC-ACGCGCGCT
 1070 1080 1090 1100 1110 1120 1130
 GGGCGTCCCACAGGAAACGCTGCCAGGATCACTGTGACTAACCGGGCGCGGTTCTGTGTCAACGACTTGCC
 1080 1090 1100 1110 1120 1130 1140
 GGCGGGC---CTGGCCAGGTCCCTGGCGTGGATCCGGC-CACGCCGATCG--TC-CTGGCGCCGGGT--CC
 1140 1150 1160 1170 1180 1190 1200
 GTGCATGACGC--CATCTGGATCCCTGAC-GCAGCGGCCACTATTCCCCGTCAAGATAACCGAACGATGAAG
 1150 1160 1170 1180 1190 1200
 G---CTGGCGCCGCAAC-GGATGAAATGGAGCTGGCCCAAT--GGCTGCGGGG---CGCCCG--GCAG
 1210 1220 1230 1240 1250 1260 1270
 TCGCGCATCGATGAT--AGGCATUTTCAATGATCAGG--GCTGCCACCTCCAAAGCCG--GTGGCCACC
 1210 1220 1230 1240 1250 1260 1270
 -CG-TCATCCCGCGATCPAAGGCAGC-GCAGGGCGCTGGCAGCGGCCACCGGCCAGGTGCGATGTGGCGCG
 1280 1290 1300 1310 1320 X
 CCTGTCGATACTCTTGAGGCCAGCGTAACTGACCGTAC--TTTCGTGAACTCAG

- *
 -
 -
 -
 -
 -
 -
 - *
 0-----
 SCORE 0: 1125 50 76 101 126 151 177 202 227
 STDEV -1 9

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	6		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	2.65

Times:	CPU	Total Elapsed
	00:03:03.09	00:09:14.00

Number of residues:	4893063
Number of sequences searched:	17731
Number of scores above cutoff:	3826

Cut-off raised to 7.
 Cut-off raised to 8.
 Cut-off raised to 9.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt.					
		Length	Score	Score	Sig.	Frame	
**** 83 standard deviations above mean ****							
1. A28214	Phosphotriesterase - Pseudomon	325	227	251	83.28	0	
**** 4 standard deviations above mean ****							
2. VGBEGX	Secreted glycoprotein gX - Pse	498	19	47	4.90	0	
3. SYBSYX	Tyrosine-tRNA ligase - Bacill	419	19	43	4.90	0	
4. NOHUG	Enolase gamma - Human #EC-numb	434	18	62	4.52	0	
5. A27124	H+-translocating ATPase - Leish	974	18	68	4.52	0	
6. S02077	Enolase gamma - Human (fragmen	433	18	62	4.52	0	
7. A24742	Enolase gamma chain - Rat #EC-	434	18	63	4.52	0	
8. A244405	Ice nucleation protein - Pseud	1200	17	63	4.15	0	
9. QDBP4L	Hypothetical protein D-206 - B	206	17	36	4.15	0	
10. A28852	Histone H3(1) - Tetrahymena py	135	17	27	4.15	0	
11. B24255	Chorion class A protein L12 pr	132	17	28	4.15	0	

**** 3 standard deviations above mean ****							
12. S01921	Hypothetical protein 1 - Chlam	451	16	19	3.77	0	
13. S01022	Hypothetical protein P-2 - Chl	86	16	19	3.77	0	
14. HSXL32	Histone H3.2 - African clawed	135	16	30	3.77	0	
15. S03605	Surface glycoprotein CD14 prec	366	16	61	3.77	0	
16. A24225	Transducin beta chain - Bovine	340	16	40	3.77	0	
17. A24853	Transducin beta chain, liver -	340	16	40	3.77	0	
18. A26066	Segmentation protein eve - Fru	376	16	35	3.77	0	
19. A25457	Transducin beta chain - Bovine	340	16	40	3.77	0	
20. DAHUAL	Arachidonate 5-lipoxygenase -	674	16	65	3.77	0	

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt.				Sig.	Frame
		Length	Score	Score			
**** 74 standard deviations above mean ****							
1. A28214	Phosphotriesterase - Pseudomon	325	227	251	74.70	0	
**** 5 standard deviations above mean ****							
2. WMBEBH	72K protein - Bovine herpesvir	664	11	72	5.42	0	
**** 4 standard deviations above mean ****							
3. QQBEE3	HHLF1 protein - Cytomegaloviru	788	9	70	4.64	0	
4. QQBEC3	HQRF1 protein - Cytomegaloviru	846	9	69	4.26	0	
5. A25902	65K protein antigen - Mycobact	588	15	69	4.26	0	
6. S05506	Phosphoenolpyruvate carboxylas	966	8	69	4.26	0	
7. S00893	Adenylate cyclase precursor -	1706	10	69	4.26	0	
8. S02389	Cyclolysin - <i>Bordetella pertus</i>	1706	10	69	4.26	0	
**** 3 standard deviations above mean ****							
9. A27124	H+-transporting ATPase - Leish	974	18	68	3.87	0	
10. WFHUM	Mullerian inhibiting factor pr	560	13	68	3.87	0	
11. GNNYSP	Genome polyprotein - Polioviru	2207	13	68	3.87	0	
12. S04255	Regulatory protein qa-1S - Neu	918	8	68	3.87	0	
13. GNWVWV	Genome polyprotein - West Nile	3430	10	68	3.87	0	
14. QQBE8	Hypothetical BPLF1 protein - E	3149	10	67	3.48	0	
15. B28894	Myeloperoxidase H7 - Human	830	11	67	3.48	0	
16. OKBOG	Protein kinase, cGMP-dependent	670	11	67	3.48	0	
17. DCZYPC	Pyruvate decarboxylase - Zymom	559	8	67	3.48	0	
18. S02386	cyaB protein - <i>Bordetella pert</i>	712	10	67	3.48	0	
19. VGBEPB	Glycoprotein gIII precursor -	479	9	67	3.48	0	
20. S00896	Ferredoxin--nitrite reductase	594	8	67	3.48	0	

1. LOW344-FIG1. PEP
A28214 Phosphotriesterase - *Pseudomonas diminuta* MG plasmid

ENTRY A28214 #Type Protein
TITLE Phosphotriesterase - *Pseudomonas diminuta* MG plasmid
pCMS1
SOURCE *Pseudomonas diminuta*
ACCESSION A28214
REFERENCE (Sequence translated from the DNA sequence)
#Authors McDaniel C. S., Harper L. L., Wild J. R.
#Journal J. Bacteriol. (1988) 170:2306-2311
#Title Cloning and sequencing of a plasmid-borne gene (opd)
encoding a phosphotriesterase.
#Name phosphotriesterase
esterase
COMMENT THIS SEQUENCE HAS NOT BEEN COMPARED TO THE
NUCLEOTIDE TRANSLATION.
SUMMARY #Molecular-weight 32243 #Length 325 #Checksum 539

SEQUENCE

Initial Score = 227 Optimized Score = 251 Significance = 74.70
 Residue Identity = 76% Mismatches = 260 Mismatches = 51
 Gaps = 28 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
MQTRVVVLKSAAAGTLLGGLAGCATWLDRSAGAIGSIRARPITISEAGFTLTHEDICGSSAGFLRAWPEFFG							
MQTRVVVLKSAAARTLILGGLAGCATWLDRSAGAMRSIRARPITISEAGFTLTHEDI---SAARQDSCVLGQS	X	10	20	30	40	50	60
	80	90	100	110	120	130	
SRKALAEKAVRGLR-ARAAGVR---TIVD----VSTFDIGRDVSLLAEVSRAADVHIVAATGLWFDPPPLSM							
SSVAQSSSGKGCGERIARQDSGWRANDCPCVDFRYRSRRQFIGR---GFAGCRR---SYLAATGLWFDPPPLSM	70	80	90	100	110	120	130
	140	150	160	170	180	190	200
RLRYVEELT----QFFLREIQYQGIEDTGIRAGIIKVATTGKATPFQELVLKAAARASLATGVPVTTHAAS							
RLRYVEELTLVLPAVRFNMASKY---TGIRAGIIKVATTGKATPFQELVLKAAARASLATGVPVTTHAAS	140	150	160	170	180	190	200
	210	220	230	240	250	260	270
QRDGGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWQ							
QRDGGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWQ	210	220	230	240	250	260	270
	280	290	300	310	320	X	330
TRALLIKALIDQGYMKQILVSNDWLFGFSSYVTNIMDVMDRVNPDGMAFIPLRVIPIFYERR							
TRALLIKALIDQGYMKQILVSNDWLFGFSSYVTNIMDVMDRVNPDGMAFIH	280	290	300	310	320	X	

2. LOW344-FIG1. PEP

WMBEBH 72K protein - Bovine herpesvirus (type 2, strain

ENTRY WMBEBH #Type Protein
 TITLE 72K protein - Bovine herpesvirus (type 2, strain
 BMV)
 DATE 31-Mar-1990 #Sequence 31-Mar-1990 #Text 31-Mar-1990
 PLACEMENT 1386.0 1.0 2.0 1.0 1.0
 SOURCE bovine mamillitis virus, bovine herpesvirus 2
 ACCESSION B29242
 REFERENCE (Sequence translated from the DNA sequence)
 #Authors Hammerschmidt W., Conraths F., Mankertz J., Pauli
 G., Ludwig H., Buhk H. J.
 #Journal Virology (1968) 165:388-405
 #Title Conservation of a gene cluster including
 glycoprotein B in bovine herpesvirus type 2
 (BHV-2) and herpes simplex virus type 1 (HSV-1).
 #Comment The amino acid sequence is not given in this paper.
 COMMENT The DNA sequence was obtained from GenBank, release
 61.0.
 COMMENT This virus is a member of the family Herpesviridae.
 SUPERFAMILY #Name herpesvirus infected cell protein ICP18.5
 SUMMARY #Molecular-Weight 72367 #Length 664 #Checksum 8190
 SEQUENCE

Initial Score = 11 Optimized Score = 72 Significance = 5.42
 Residue Identity = 26% Mismatches = 101 Mismatches = 203
 Gaps = 83 Conservative Substitutions = 0

X	10	20	30	40	50	
MOTPRVVLKSAAGSTILL	GGL	AGCATWLDRSAQAI	SIRARPITISEA	GFTLTHEDI		
RLAGKICDHVTBQARVRL	DADEMRRNLPHV	VGLSEARRARALHA	LEVSSKMTEANSG	GPAAQAE		
230 X 240	250	260	270	280	290	
60	70	80	90	100	110	120
CGSSAGFLRAWPEFFGSRK	ALAEKAVRGLR	ARAAGVRTIVDV	STFDIGRD	VSLLAEVSR	---AADV	HIV
REASA	LLDAHHVFKSAPPGL	YAVSEI	RFWLSSGDR	SGSTVDAFADNL	SALAER	RERRYETGAV
300	310	320	330	340	350	360
130	140	150	160	170	180	
AATG	---LWFDPPLSMR	---I	RYVEEL	TQFFLREI	QYGIEDT	GIRAGI
400	410	420	430	440	450	460
190	200	210	220	230	240	
--KAAARASLATGV	PVTHTA	ASWRDG	---ERGRPPFLSPK	LEPSRVCIGH	SDDTDDLSYLT	TALLRGYLI
470	480	490	500	510	520	530
250	260	270	280	290	300	310
GLDHIPHSAIGLE	DNASASPLLGIR	SWOTRALLIKALID	QGY	---MKQILVSN	--DWLFGFSSY	VTNIMD
540	550	560	570	580		
320	330	340	X			
VMD	-RVNPDGMAFIPLRV	IPFYERRAS	-HPKPCGASL			
500	510	520	530	540		
VIDNRGNPD	--TFDTHR	---FMRASLMRHRV	DFAALLPGITHQFFE			

3. LOW344-FIG1. PEP

QQBEE3 HHLF1 protein - Cytomegalovirus, (strain AD169)

ENTRY	QQBEE3	#Type Protein
TITLE	HHLF1 protein - Cytomegalovirus (strain AD169)	
DATE	30-Sep-1989 #Sequence 30-Sep-1989 #Text 31-Dec-1989	
PLACEMENT	1358.0	1.0 1.0 2.0 1.0
SOURCE	human cytomegalovirus, human herpesvirus 5	
ACCESSION	C27349	
REFERENCE	(Sequence translated from the DNA sequence)	
#Authors	Weston K., Barrell B. G.	
#Journal	J. Mol. Biol. (1986) 192:177-208	
#Title	Sequence of the short unique region, short repeats, and part of the long repeats of human cytomegalovirus.	
COMMENT	The DNA sequence was obtained from EMBL, release 13.	
COMMENT	This virus is a member of the family Herpesviridae.	
GENETIC		
#Name	HHLF1	
SUPERFAMILY	@Name Cytomegalovirus HBLF1 protein	
SUMMARY	#Molecular-weight 83881 #Length 788 #Checksum 7858	
SEQUENCE		

Initial Score	=	9	Optimized Score	=	70	Significance	=	4.64
Residue Identity	=	23%	Matches	=	90	Mismatches	=	241
Gaps	=	56	Conservative Substitutions	=			=	0

X	10	20	30	40	50
MOTPRVVLKSAAGSTILL	GGL	AGCATWLDRSAQAI	SIRARPITISEA	GFTLTHEDI	

4. LOW344-FIG1. PEP

QBEC3 HQRF1 protein - Cytomegalovirus (strain AD169)

ENTRY QQBEC3 #Type Protein
 TITLE HQRF1 protein - Cytomegalovirus (strain AD169)
 DATE 30-Sep-1989 #Sequence 30-Sep-1989 #Text 30-Sep-1989
 PLACEMENT 1358.0 1.0 1.0 1.0 1.0
 SOURCE human cytomegalovirus, human herpesvirus 5
 ACCESSION C26078
 REFERENCE (Sequence translated from the DNA sequence)
 #Authors Weston K., Barrell B. G.
 #Journal J. Mol. Biol. (1986) 192:177-208
 #Title Sequence of the short unique region, short repeats,
 and part of the long repeats of human
 cytomegalovirus.
 COMMENT The DNA sequence was obtained from EMBL, release 13.
 COMMENT This virus is a member of the family Herpesviridae.
 GENETIC
 #Name HQRF1
 SUPERFAMILY #Name cytomegalovirus HQRF1 protein
 SUMMARY #Molecular-weight 91047 #Length 846 #Checksum 2604
 SEQUENCE

Initial Score = 9 Optimized Score = 69 Significance = 4.2
Residue Identity = 22% Matches = 86 Mismatches = 24
Gaps = 46 Conservative Substitutions =

X 10 20 30 40 50
M**Q**TRRVVVLKSAAGTLLGGLAGCATWLDRSA-QAIG-----SIRARPITISEAGFTLTHED

ASAPHPASLITAVRRHLNRRICCGWLAI GAVI PARWI GCAAGPATGTAAGTISPPAASGTEAAGGDAPCA

330	X	340	350	360	370	380	390
60	70	80	90	100	110	120	
ICG--SSAGFLRAWPEFFGSRK---LA-EKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVH							
400	410	420	430	440	450	460	
130	140	150	160	170	180		
IVAATGLWF-DPPLSMRLRYVEELTQFFLREIQQYGIEDTG-IRAGIIKVATTGKATPFQE-----LVLKA							
470	480	490	500	510	520	530	
190	200	210	220	230	240	250	
AARASLATGVPVTTHTAASQRDGE-RGRPPFLSPKLEPSRVCIGHSDTDDLSYLTALLRGYLIGLDHIPHS							
540	550	560	570	580	590	600	
260	270	280	290	300	310		
AIGLEDNASASPLLGIRSWQT---RALLIKALIDQGQYMKQILVSNWLFGFSSYV-----TNIMDVMDRVNP							
610	620	630	640	650	660		
320	330	340	X				
D-GMAFIPLR-VIPFYERRASHRKRCQASL							
670	680	690	700				

5. LOW344-FIG1. PEP

A25902 65K protein antigen - *Mycobacterium leprae*

ENTRY A25902 #Type Protein
 TITLE 65K protein antigen - *Mycobacterium leprae*
 SOURCE *Mycobacterium leprae*
 ACCESSION A25902
 REFERENCE (Sequence translated from the DNA sequence)
 #Authors Mehra V., Sweetser D., Young R. A.
 #Journal Proc. Nat. Acad. Sci. USA (1986) 83:7013-7017
 #Title Efficient mapping of protein antigenic determinants.
 SUMMARY #Molecular-weight 61855 #Length 588 #Checksum 3048
 SEQUENCE

Initial Score = 15 Optimized Score = 69 Significance = 4.26
 Residue Identity = 23% Matches = 90 Mismatches = 231
 Gaps = 58 Conservative Substitutions = 0

X	10	20	30	40	50	
M@TRRVVLKSA---A@TLLNGLAGCATWLDRSAQAISSIR---ARPITI---SEAGFTLTHED						
10	20	30	40	50	60	
VPGRDGET@PASCGRFSRALHPA@V@.GCRHPVT-----LASFLIRRNFAMAKTIAYDEEARRGL--ER						

60	70	80	90	100	110	120	
ICGSAGFLRAWPEFFGERKAIAEKAVRGLRARAAGVRTIVDVSTFD----IGRDVSLLAEVSRAADVHIVA							
70	80	90	100	110	120	130	
130	140	150	160	170	180	190	
ATGLWFDPPLSMPLRYVEELTQFFLREIQQYGIEDTG@PAGI----IKVATT-GKATPFQELVLKAAARASLA							

GDGTTTATVLADAL---VREH----PQV/RS25-MAFL&LKRLEKAVDKVTETLLKDAKEVETKERBIAATAAIS
 140 150 160 170 180 190
 200 210 220 230 240 250
 TGVPVTTHTAASARD---GERIE----RPFPLSPKLM--EPSRVCIG--HSDDTDLDSYL TALLRGYLI GLDH
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 AGDQSIGDLIAEANDKVGNEGVITVIEESNIFGLGLELREGMRFDKGYISGYFVTDAERQEAVLLEPYILLVS
 200 210 220 230 240 250 260
 260 270 280 290 300 310 320
 IPHSAIGLEDNASASPLLGIRESWOTTPALLUKALIDQGYMKQILVSNDWLFGFSSYVTNIMDVMDR--VNPDG
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 SKVS----TVKDILLPLLEKVI BARKSLLIIAEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQD
 270 280 290 300 310 320 330
 330 340 X
 MAFIPLRVIPFYERRASHRKRCBASL
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 MA-ILTGAQVISEEVLTLLENTDLALLLQKARKVVMT
 340 350 360 370

6. LOW344-FIG1. PEP

S05506 Phosphoenolpyruvate carboxylase 1 - Common ice pla

ENTRY S05506 #Type Protein
 TITLE Phosphoenolpyruvate carboxylase 1 - Common ice plant
 #EC-number 4.1.1.31
 SOURCE Mesembryanthemum crystallinum #Common-name common
 ice plant.
 ACCESSION S05506
 REFERENCE
 #Authors Cushman J. C., Bohnert H. J.
 #Journal Nucleic Acids Res. (1983) 17:6745
 #Title Nucleotide sequence of the gene encoding a CAM
 specific isoform of phosphoenolpyruvate
 carboxylase from Mesembryanthemum crystallinum.
 #Reference-number S05506
 #Accession S05506
 #Molecule-type DNA
 #Residues 1-966 (CDS)
 REFERENCE
 #Authors Pickers J., Cushman J. C., Michalowski C. B., Schmitt
 J. M., Bohnert H. J.
 #Journal Mol. Gen. Genet. (1989) 215:447-454
 #Title Expression of the CAM-form of phosphoenolpyruvate
 carboxylase and nucleotide sequence of a full
 length cDNA from Mesembryanthemum crystallinum.
 #Reference-number S02716
 #Accession S02716
 #Molecule-type mRNA
 #Residues 1-966 (RIC)
 SUMMARY #Molecular-weight 110659 #Length 966 #Checksum 725
 SEQUENCE

Initial Score = 8 Optimized Score = 69 Significance = 4.26
Residue Identity = 23% Matches = 94 Mismatches = 212
Gaps = 50 Conservative Substitutions = 0

X	10	20	30	40	50		
M@TPRVVVLKSARAGTILL@P@LACCATWLDR@SABA---IGSIR---ARPITISEAGFTLTHEDI							
SVRRSLLQKHGRIDOLARLYAKD@TPD@KELDEAL@REIQA@AFRTDEIRRT@PTP@DEMRA@GMSYFHETI	180	190	200	210	220	230	240
60	70	80	90	100	110		

CESSAGFLRAWPEFFGSRKALAEKAVRLRPR-----AAGVRTIVDVSTFDIGRDVSLLAEVSR
 WNGVPKFLR-----RLDTALK-NIGITERVFNAPLIQFSSWMGGDRDGNPRVTPEVTRDVCLLA-RMM
 250 260 270 280 290 300 310
 120 130 140 150 160 170
 AADVHIVAATGLWFDPPLSM----PLR-YVEELTQFFLREIQYGIEDTGIRAGIIKVATTGKATPFQE--L
 111 111 111 111 111 111 111
 AANMYFSQIDELMF--ELSMWPCDIELRERAEELHKYSKRDSDKHYIE-----FWKQIPSSEPYR
 320 330 340 350 360
 180 190 200 210 220 230
 VLKA-----AARASLATGV---PV-TTHTAASQRDGERGRPPFLSP-KLEPSRVCIGHSDDTDDLS
 111 111 111 111 111 111 111
 VILADVRDKLYYTRERSRQLLASEVSEIPVEATFTETDQ-----FLEPLECYRSLCACGDRPVADGS
 370 380 390 400 410 420 430
 240 250 260 270 280 290 300
 YL-----TALLRGYLIGLDHIPHSAGILEDNASASPLLGIRSWQTRALLIKALIDQGYMKQILVSNNDWLFGF
 111 111 111 111 111 111 111
 LLDFMQRQVATFGLCLVKLDIRESERHTDVMDAITTHLGIGS--YRDWTEEKRQD--WLLSELRGKRPLFGP
 440 450 460 470 480 490
 310 320 330 340 X
 SSYVT-NIMDVMDRVNPDGMAFIPLFV1FFY-ERRAGHRKRCQASL
 111 111 111 111 111 111 111
 DLPRTDEIADVLDTIN--VIAELPQDSFGAYVISMATAPSQVLAPELLQRECKVKK
 500 510 520 530 540 X 550

7. LOW344-FIG1. PEP

500893 Adenylate cyclase precursor - *Bordetella pertussis*

ENTRY 500893 #Type Protein
 TITLE Adenylate cyclase precursor - *Bordetella pertussis*
 #EC-number 4.6.1.1
 INCLUDES probable haemolysin
 SOURCE *Bordetella pertussis*
 ACCESSION 500893
 REFERENCE
 #Authors Glaser P., Ledant D., Sezer O., Pichot F., Ullmann
 A., Danchin A.
 #Journal Mol. Microbiol. (1993) 2:19-30
 #Title The calmodulin-sensitive adenylate cyclase of
Bordetella pertussis: cloning and expression in
Escherichia coli.
 #Reference-number 500893
 #Accession 500893
 #Molecule-type DNA
 #Residues 1-1706 (GLA)
 #Cross-reference EMBL:Y00545
 GENETIC
 #Name cya
 SUMMARY #Molecular-weight 177506 #Length 1706 #Checksum 6461
 SEQUENCE

Initial Score = 10 Optimized Score = 69 Significance = 4.26
 Residue Identity = 23% Matches = 93 Mismatches = 206
 Gaps = 97 Conservative Substitutions = 0

X 10 20 30 40
 MQTRRVVLKSAAAGTILLIGGLAGCATWLDRSAQAIGGSIR-----ARPITI---SEA
 111 111 111 111 111 111 111 111
 LMTQFGRAGSTNTPDEAASLAAAVHGL-GEASSAVAETVSGFFRGSSRWAGGGFGVAGGGAMALGGGIAAAVGA
 490 500 510 520 530 540 550

50	60	70	80	90	100	110
GFTLTHEDICG--SSAGFLRAWFEEFG----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAE						
100	110	120	130	140	150	160
110	120	120	130	140	150	160
GMSLTDAPAGQKAAAGAEIALQLTGTIVELASSIAAL-----LAAARGVTSGLQVAGASAG-----						
160	170	180	190	200	210	220
170	180	180	190	200	210	220
VSRAADVHIVAATGLWFDPPLSMRI.RYVEELTQFFLREIQYGIE-----DTGIRAGIIKVATT						
220	230	240	250	260	270	280
230	240	240	250	260	270	280
---AAAGALAAALSPMEIYGLVQDQSHYADQLDKLAQESSAYGYEGDALLAQLYRDKTAAEGAVAGVSAVLST						
280	290	300	310	320	330	340
290	300	300	310	320	330	340
610 620 630 640 650 660 670						
340	350	360	370	380	390	400
350	360	360	370	380	390	400
GKATPFQELVLKAAARASLATGVPVTTHTAASBRDG----ERGRPPFLSPKLEPSRVCIGHSSDDTDDLSYL						
400	410	420	430	440	450	460
410	420	420	430	440	450	460
VGA----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGVQQPIIEKL----ANDYARKIDELGGP						
460	470	480	490	500	510	520
470	480	480	490	500	510	520
680 690 700 710 720 730						
520	530	540	550	560	570	580
530	540	540	550	560	570	580
TA----LLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWATRALLIKALIDQGYMKQILVSNDWLFGF						
580	590	600	610	620	630	640
590	600	600	610	620	630	640
QAYFEKNLQARHEQLANSQGILRKMLADLQAGWNASSVIG---VQTTEISKSAL----ELAAITGNADNL--K						
640	650	660	670	680	690	700
650	660	660	670	680	690	700
740 750 760 770 780 790						
700	710	720	730	740	750	760
710	720	720	730	740	750	760
310 320 330 340 X						
SSYVTNIMDVMDRVNPQGMAFIP--LRVIPFYERRASHRKRCQASL						
760	770	780	790	800	810	820
770	780	780	790	800	810	820
SVDV----FVDRFVQGERVAGQPVVLDVAAGGIDIAS-RKGERPALTFITPLAAPG						
820	830	840	850	860	870	880
830	840	840	850	860	870	880

8. LDW344-FIG1. PEP

S02389 **Cyclolysin - *Bordetella pertussis***

Initial Score = 10 Optimized Score = 69 Significance = 4.26
Residue Identity = 23% Matches = 93 Mismatches = 206
Gaps = 97 Conservative Substitutions = 0

X 10 20 30 40
M**ATP**RVV**U**L**K**SAAAG**T**LL**R**GL**A**GG**C**AT**W**L**D**RS**A**QA**G**GS**I**R-----**A**RP**I**T**I**---**S**E**A**

LMTQFGRAGSTNTPQEAAASL8AAMVPL-GEASRVAETVSGFFRGSSRWAGGGFGVAGGGAMALGGGIAAAVGA
 490 500 510 520 530 540 550
 50 60 70 80 90 100 110
 GFTLTHEDICG--SSAQFLRAWPEIFG-----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAE
 560 570 580 590 600
 GMSLTDDAPAGBKAAGAEIALBLTGSTVELASSIALA-----LAAARGVTSGLQVAGASAG-----
 120 130 140 150 160 170
 VSRAADVHIVAATGLWFDPLSMRLRYVEEL1QFFLREIQYGLIE-----DTGIRAGIIVKATT
 610 620 630 640 650 660 670
 ---AAAGALAAAALSPMEIYGLVQQSHYAJQALDKLAQESSAYGYEGDALLAQLYRDKTAAEGAVAGVSAVLST
 180 190 200 210 220 230
 GKATPFQELVLKAAPRASLATGVPVTTHTAASQRDG-----ERGRPPFLSPKLEPSRVCIGHSDDTDDLSYL
 680 690 700 710 720 730
 VGA-----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGVQQPIIEKL-----ANDYARKIDELEGGP
 240 250 260 270 280 290 300
 TA-----LLRGYLIGLDHIPHSAGLEDNASASPLLGIRSWOTRALLIKALIDQGYMKQILVSNDWLFGF
 740 750 760 770 780 790
 QAYFEKNLQARHEQLANSQGLRKMILADLQAGWNASSVIG---VQTEIISKSAL---ELAAITGNADNL--K
 310 320 330 340 X
 SSYVTNIMDVMDRVNPDGMAFIP--LRVIPFYERRASHRKRCBASL
 800 810 820 830 840
 SVDV-----FVDRFVQGERVARGPVPVLDVAAGGIDIAS-RKGERPALTFTITPLAAPG

9. LOW344-FIG1. PEP

A27124 H+-transporting ATPase - Leishmania donovani

ENTRY A27124 #Type Protein
 TITLE H+-transporting ATPase - Leishmania donovani
 #EC-number 3.6.1.35
 ALTERNATE-NAME proton-transporting ATPase
 SOURCE Leishmania donovani
 ACCESSION A27124
 REFERENCE (Sequence translated from the DNA sequence)
 #Authors Meade J. C., Shaw J., Lemaster S., Gallagher G.,
 Stringer J. R.
 #Journal Mol. Cell. Biol. (1987) 7:3937-3946
 #Title Structure and expression of a tandem gene pair in
 Leishmania donovani that encodes a protein
 structurally homologous to eucaryotic
 cation-transporting ATPases.
 #Comment The authors translated the codon AGA for residue 352
 as Lys.
 SUMMARY #Molecular-weight 107476 #Length 974 #Checksum 834
 SEQUENCE

Initial Score = 13 Optimized Score = 68 Significance = 3.87
 Residue Identity = 24% Matches = 93 Mismatches = 210
 Gaps = 82 Conservative Substitutions = 0

X 10 20 30 40 50
 MQTRR-----VVLKSAANGTLLAGLAGCATWLDRSAQAIGSIRARPITISEAGFTLTHE---
 500 510 520 530 540 550 560
 FLDPPRPDTKD1TIRRSKEYGVDVKHNITGQHLLIAKEMC-RMLDLDPNILTADKLPQIKDANDLPEDLGEKYG

60	70	80	90	100	110	120
DICGSSAGIFLRAWPEFFGSRKALAIKAVIPGLRAR---ANGVRTIVDVSTFDIGRDVSLAEVSRAADVHIVA						
570	580	590	600	610	620	630
130	140	150	160	170		
ATGLWFDPPPLSMRRLRYVE---ELTQFPLRGIQYGIEDT---GIRAGIIVKATTGKA---TPFQ---E						
ADMVLTEPGLS---VVVEAMLVSREVFGIRMLSFLTYRISATLQLVCFFFIAACFSLTPKAYGSVDPHFQFFFHL						
640	650	660	670	680	690	700
180	190	200	210	220	230	240
LVLKAAARASLATGVPVT---THTAASVURDGERGRFPFLSPKLEPSRVCIGHSDTDDLSYLTALLRGYLIG						
710	720	730	740			
PVLMFMLITLLNDGCLMTIGYDHVTPS---ERPQKWNL-PVVFVS-----ASILAAVACGSSLM						
250	260	270	280	290	300	310
LDHIPHSAGILEDNASASPLLGIRSWBTR---ALLIKALIDQGYMKQILVSNDLFGFSSYVTNIMDVMDR						
760	770	780	790	800	810	
320	330	340	X			
VNP--DGMAFIPLRVIPFYERRAS--HRKRCQASL						
820	830	840	850	860		
PSPILFCGAIISLLV---STMAASFWHKERPDNVLTGGLAWGQTN						

10. LOW344-FIG1. PEP

WFHUM Mullerian inhibiting factor precursor - Human

ENTRY WFHUM #Type Protein
 TITLE Mullerian inhibiting factor precursor - Human
 ALTERNATE-NAME Mullerian inhibiting substance (MIS)
 DATE 13-Aug-1986 #Sequence 13-Aug-1986 #Text 30-Jun-1987
 PLACEMENT 596.0 5.0 1.0 1.0 1.0
 SOURCE Homo sapiens #Common-name man
 ACCESSION A01397
 REFERENCE (Sequence translated from the DNA sequence)
 #Authors Cate R. L. , Mattaliano R. J. , Hession C. , Tizard R. ,
 Farber N. M. , Cheung A. , Ninfa E. G. , Frey A. Z. ,
 Gash D. J. , Chow E. P. , Fisher R. A. , Bertomis J. M. ,
 Torres G. , Wallner B. P. , Ramachandran K. L. , Ragin
 R. C. , Mangano T. F. , MacLaughlin D. T. , Donahoe
 P. K.
 #Journal Cell (1986) 45:685-698
 COMMENT This protein is homologous to transforming growth
 factor beta, inhibin alpha chain, and inhibin A
 and B chains. The area of best homology
 corresponds to the mature proteins.
 COMMENT Although it does not compete with EGF for receptor
 binding sites, MIS can inhibit the
 autophosphorylation of the EGF receptor in vitro.
 GENETIC
 #Introns 138/1, 185/3, 222/1, 275/2
 SUPERFAMILY #Name inhibin
 KEYWORDS testicular glycoprotein\ gonadal differentiation\
 antitumor agent\ Mullerian duct\ TGF-beta homolog\
 inhibin homolog
 FEATURE
 1-25 #Domain signal and propeptide sequence
 (SIGNAL)

26-560

#Protein Mullerian inhibiting factor
(MIF)
(MAT)

64,329

#Binding-site carbohydrate (Asn)
(potential)SUMMARY
SEQUENCE

#Molecular-weight 59192 #Length 560 #Checksum 3812

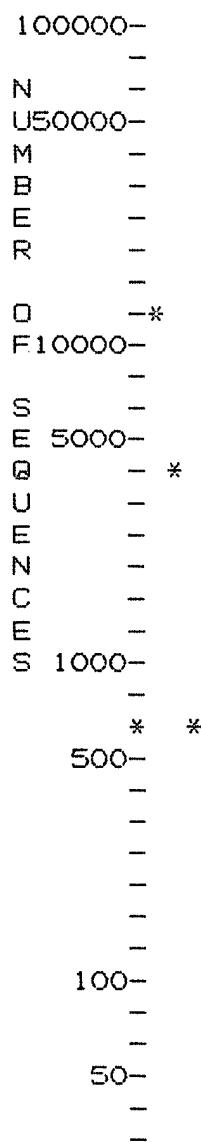
Initial Score = 13 Optimized Score = 68 Significance = 3.87
 Residue Identity = 24% Matches = 99 Mismatches = 211
 Gaps = 92 Conservative Substitutions = 0

	10	20	30	40	50		
MQTRRVVVL-KSAAAGTLLG-GLA-----GCATWL-DRSAQAA-IGSIRARPIT-ISEAGFTL							
LPGAQSLCPSPRDTRYLVLAVDRPAGAWRGSGLALTLQPRGEDSRLSTARLQALLFGDDHRCFTRMTPALLL	190	200	210	220	230	240	250
60 70 80 90 100 110							
THEDICGSSA-GFLRAWPEFFGSRKALAE---KAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSR-AA							
PRSEPAAPLPAHQLDTVPFPRPPRPSAELEESPPSADPFLETLTRLVR-----ALRVPPARASAPRLAL	260	270	280	290		300	310
120 130 140 150 160 170 180							
DVHIVA--ATGL--WFDPPLSMRLRYVEELTQFFLREIQYGYIEDTGIRAGIIVKATTGKATPFQELV---LK							
DPDALAGFPQGLVNLSDPAALERLILDGEEEPLLLRLR---PTAATTGDPAPLHDPTSAPWATALARRVAELQ	320	330	340	350	360	370	380
190 200 210 220 230 240							
AAA--RASLATGVPVTTHTAASQRDGERGRPPFLSPKLEPSRVCIGHSDDDLSYLTAL--LRGYLIGLD-							
AAAAELRSLPGLPPATAPLLA-----RLLALCP-GGP-----GGLGDPLRALLLKLQGLRVEWRGRDP	390	400	410		420	430	440
250 260 270 280 290 300 310							
HIPHSA-IGLEDNASASP-LLGIRSWETRA---LLI---KALIDQGYMKQILVSNDLFGFSSYVTNIMDV							
RGPGRAGRSAGATAADGPCALRELGVOLRAERSVLIPETYQANNCGG-----VCGWPQSDRNPYRG	450	460	470	480	490		
320 330 340 X							
MDRVNPQGM-----AFIPLRVIPFY-----ER-RASHRKRCQASL							
NHVVLLLKMQARGAALARPPCCVPTAYAGKILISLSEERISAHVPMVATECGCR	510	520	530	540	550	X	560

Results file low344-fig1-spt.res made by maryh on Wed 17 Apr 91 11:34:12-PDT.

Query sequence being compared: LOW344-FIG1.PEP
Number of sequences searched: 15409
Number of scores above cutoff: 3758

Results of the initial comparison of LOW344-FIG1.PEP with:
Data bank : Swiss-Prot 14, all entries



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	6		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

SEARCH STATISTICS

Scores: Mean 6 Median 7 Standard Deviation 2.69

Times:	CPU	Total Elapsed
	00:03:15.00	00:13:53.00

Number of residues: 4914263
Number of sequences searched: 15409
Number of scores above cutoff: 3758

Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init.	Opt.	Score	Score	Sig.	Frame
**** 82 standard deviations above mean ****								
1. OPD\$PSEDI	PHOSPHOTRIESTERASE (EC 3.6. -.-)	325	227	251	82.28		0	
**** 4 standard deviations above mean ****								
2. MTC3\$CHVN1	MODIFICATION METHYLASE CVIB II	377	19	25	4.84		0	
3. SYY\$BACCA	TYROSYL-TRNA SYNTHETASE (EC 6.1.4.1)	419	19	43	4.84		0	
4. VGLX\$PRV	SECRETED GLYCOPROTEIN GX (GENE)	498	19	47	4.84		0	
5. ENOG\$RAT	GAMMA ENOLASE (EC 4.2.1.11) (?)	433	18	63	4.47		0	

6.	ENOG\$HUMAN	GAMMA ENOLASE (EC 4.2.1.11) (2	433	18	62	4.47	0
7.	ATXA\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	4.47	0
8.	ICEN\$ERWHE	ICE NUCLEATION PROTEIN (GENE N	1258	18	63	4.47	0
9.	ATXB\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	4.47	0
10.	H31\$TETPY	HISTONE H3. 1.	135	17	27	4.10	0
11.	KCCA\$RAT	CALCIUM/CALMODULIN-DEPENDENT P	478	17	57	4.10	0
12.	KCCA\$MOUSE	CALCIUM/CALMODULIN-DEPENDENT P	478	17	57	4.10	0
13.	ICEN\$PSESY	ICE NUCLEATION PROTEIN (GENE N	1200	17	63	4.10	0
14.	CHA2\$BOMMO	CHORION CLASS A PROTEIN L12 PR	132	17	28	4.10	0
15.	Y206\$LAMBD	HYPOTHETICAL PROTEIN ORF206.	206	17	36	4.10	0
**** 3 standard deviations above mean ****							
16.	ARSA\$HUMAN	ARYLSULFATASE A PRECURSOR (EC	507	16	64	3.72	0
17.	HMEV\$DROME	SEGMENTATION PROTEIN EVEN-SKIP	376	16	35	3.72	0
18.	GBB1\$HUMAN	GUANINE NUCLEOTIDE-BINDING PRO	340	16	40	3.72	0
19.	H32\$XENLA	HISTONE H3. 2.	135	16	30	3.72	0
20.	CD14\$MOUSE	CD14 DIFFERENTIATION ANTIGEN P	366	16	61	3.72	0

The scores below are sorted by optimized score.
 Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init.	Opt.	Score	Score	Sig.	Frame
**** 76 standard deviations above mean ****								
1. OPD\$PSEDI	PHOSPHOTRIESTERASE (EC 3.5. -.-)	325	227	251	76.20	0		
**** 4 standard deviations above mean ****								
2. YHL1\$HCMVA	HYPOTHETICAL PROTEIN HHLF1.	788	9	70	4.74	0		
3. YQR1\$HCMVA	HYPOTHETICAL PROTEIN HQRF1.	846	9	69	4.34	0		
4. CYAA\$BORPE	CALMODULIN-SENSITIVE ADENYLATE	1706	10	69	4.34	0		
5. CAP1\$MESCR	PHOSPHOENOLPYRUVATE CARBOXYLAS	966	8	69	4.34	0		
**** 3 standard deviations above mean ****								
6. MIS\$HUMAN	MULLERIAN INHIBITING FACTOR PR	560	13	68	3.95	0		
7. ATXA\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	3.95	0		
8. POLG\$WNV	GENOME POLYPROTEIN (CAPSID PRO	3430	10	68	3.95	0		
9. ATXB\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	3.95	0		
10. ATPO\$OENBI	ATP SYNTHASE ALPHA CHAIN, MITO	511	9	68	3.95	0		
11. POLG\$POL2L	GENOME POLYPROTEIN (COAT PROTE	2207	13	68	3.95	0		
12. PGKH\$WHEAT	PHOSPHOGLYCERATE KINASE, CHLOR	480	8	68	3.95	0		
13. EXON\$HSV11	ALKALINE EXONUCLEASE (EC 3.1.1)	626	9	67	3.55	0		
14. KGP\$BOVIN	CGMP-DEPENDENT PROTEIN KINASE	670	11	67	3.55	0		
15. KGPB\$HUMAN	CGMP-DEPENDENT PROTEIN KINASE,	686	11	67	3.55	0		
16. VGL3\$PRV	GLYCOPROTEIN GIII PRECURSOR	479	9	67	3.55	0		
17. PHYB\$ARATH	PHYTOCHROME B (GENE NAME: PHYB	1172	9	67	3.55	0		
18. ATI1\$HSV11	ALPHA TRANS-INDUCING FACTOR 73	693	9	67	3.55	0		
19. PGCA\$RAT	CARTILAGE-SPECIFIC PROTEOGLYCA	2124	9	67	3.55	0		
20. PYR1\$YEAST	CARBAMOYL-PHOSPHATE SYNTHETASE	1456	11	67	3.55	0		

1. LOW344-FIG1. PEP

OPD\$PSEDI PHOSPHOTRIESTERASE (EC 3.5. -.-) (GENE NAME: OPD).

ID	OPD\$PSEDI	STANDARD:	PRT:	325 AA.
AC	P13739;			
DT	01-JAN-1990	(REL. 13, CREATED)		
DT	01-JAN-1990	(REL. 13, LAST SEQUENCE UPDATE)		
DT	01-APR-1990	(REL. 14, LAST ANNOTATION UPDATE)		
DE	PHOSPHOTRIESTERASE (EC 3.5. -.-)	(GENE NAME: OPD).		
OS	PSEUDOMONAS DIMINUTA.			
OG	PLASMID PCMS1.			
OC	PROKARYOTA; BACTERIA; GRAM-NEGATIVE AEROBIC RODS AND COCCI;			
OC	PSEUDOMONADACEAE.			

RN [1] (STRAIN MG; SEQUENCE FROM N. A.)
 RA McDANIEL C. S. , HARPER L. L. , WILD J. R. ;
 RL J. BACTERIOL. 170:2306-2311(1988).
 CC -!- PATHWAY: PESTICIDE DETOXIFICATION.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
 DR EMBL; M20392; PPPTE.
 DR PIR; A28214; A28214.
 KW PLASMID; HYDROLASE; MEMBRANE.
 SQ SEQUENCE 325 AA; 35583 MW; 528207 CN;

Initial Score = 227 Optimized Score = 251 Significance = 76.20
 Residue Identity = 76% Matches = 260 Mismatches = 51
 Gaps = 26 Conservative Substitutions = 0

X	10	20	30	40	50	60	70
MQTRRVVLKSAAAGTLLGGLAGCATVLDRSAQAIGSIRARPITISEAGFTLTHEDICGSSAGFLRAWPEFFG							
MQTRRVVLKSAAARTLLGGLAGCATVLDRSAQAMRSIRARPITISEAGFTLTHEDI---SAARQDSCVLGQS	X	10	20	30	40	50	60
	80	90	100	110	120	130	
SRKALAEKAVRGLR-ARAAGVR----TIVD----VSTFDIGRDVSSLAEVSRAADVHIVAATGLWFDPPPLSM							
SSVAQSSSGKGCGCERIARDQSGWRANDCRCVDFRYRSRRQFIGR---GFAGCRR---SYLAATGLWFDPPPLSM	70	80	90	100	110	120	130
	140	150	160	170	180	190	200
RLRYVEELT----QFFLREIQYGIEDTGIRAGI IKVATTGKATPFQELVLKAAARASLATGVPVTTHTAAS							
RLRYVEELTLVLPAVRFNMASKY----TGIRAGI IKVATTGKATPFQELVLKAAARASLATGVPVTTHTAAS	140	150	160	170	180	190	200
	210	220	230	240	250	260	270
QRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHSAGLEDNASASPLLGIRSWQ							
QRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHSAGLEDNASASPLLGIRSWQ	210	220	230	240	250	260	270
	280	290	300	310	320	X	330
TRALLIKALIDQGYMKQILVSNOWLFGFSSYVTNIMDVMDRVNPDGMAFIPLRVIPFYERR							
TRALLIKALIDQGYMKQILVSNOWLFGFSSYVTNIMDVMDRVNPDGMAFIH	280	290	300	310	320	X	

2. LOW344-FIG1. PEP

YHL1\$HCMVA HYPOTHETICAL PROTEIN HHLF1.

ID YHL1\$HCMVA STANDARD; PRT; 788 AA.
 AC P09695;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL PROTEIN HHLF1.
 OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
 OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE.
 RN [1] (SEQUENCE FROM N. A.)
 RA WESTON K. , BARRELL B. G. ;
 RL J. MOL. BIOL. 192:177-208(1986).
 CC -!- SIMILARITY: TO HHLF1, HHLF5, HHLF6, HHLF7, AND HQRF1.
 DR EMBL; X04630; HEHCMVU.
 DR PIR; C27349; QBEBE3.
 KW HYPOTHETICAL PROTEIN.
 FT CARBOHYD 76 76 POTENTIAL.
 FT CARBOHYD 110 118 POTENTIAL.
 FT CARBOHYD 223 223 POTENTIAL.

SQ SEQUENCE 788 AA; 80481 MW; 2964529 CN;
 Initial Score = 9 Optimized Score = 70 Significance = 4.74
 Residue Identity = 23% Matches = 50 Mismatches = 241
 Gaps = 50 Conservative Substitutions = 0

X	10	20	30	40	50	
M@TRRVVLKSAAGTLLGGLAGCATWLDRSA-QAIG-----SIRARPITISEAGFTLTHED						
ASAPHPASLLTAVRRHLNQRICCGWLALGAVLPARWLGCAGPATGTAAGTTSPPAASGTETEAAGGDAPCA						
330 X	340	350	360	370	380	390
60	70	80	90	100	110	120
,ICG--SSAGFLRAWPEFFGSRKA---LA-EKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVH						
IAGAVGSAVPVPIQPYGAAGGGAI CVPNADAHAVVGADAAAAAAPTVMVGSTAMAGFAAS--GTVPRAMLVV						
400	410	420	430	440	450	460
130	140	150	160	170	180	
IVAATGLWF-DPPLSMRLRYVEELTQFFLREI@YGIEDTG-IRAGI IKVATTGKATPFQE---LVLKA						
LLDELGAVFGYCPPLDGHVYPLAAEI SHFIRAGVLGALALGRESAPAAEAARRLLPELDREQWERPRWDALHL						
470	480	490	500	510	520	530
190	200	210	220	230	240	250
AARASLATGVPVTHTAASQRDGERGRPPFLSPKLEPSRVCIGHSSDDTDDLSYLTALLR--GYLIGLD--HI						
HPRAALWAREP-HGQLAFLLRPG-RGEAEVLTATKHPAICANVEDYLQD---ARRRADABALGLDLATV						
540	550	560	570	580	590	600
260	270	280	290	300	310	
PHSAIG-----LEDNASASPLLGIRSW@TRALLIKALIDQG@YMKQILVSNDWLFGFSSYVTNIMDVMD						
VMEAGGQMIHKKTKKPKGKEDESLMKGKHSRYTR-PTEPPLTPQASLGRALRRDDEDWKPS---RLPGED						
610	620	630	640	650	660	
320	330	340	X			
RVNPDGMAFIPL---RVIPFYERR-ASHRKRCQASL						
SWYDLDDETFWVLGSNRKNDVYQRRWKKTVLRCGLEIDRPMPTVPKG						
670	680	690	700	X	710	

3. LOW344-FIG1. PEP
YQR1\$HCMVA HYPOTHETICAL PROTEIN H0RF1.

ID YQR1\$HCMVA STANDARD: PRT: 846 AA.
 AC P09715;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL PROTEIN H0RF1.
 OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
 OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE.
 RN [1] (SEQUENCE FROM N. A.)
 RA WESTON K. , BARRELL B. G. ;
 RL J. MOL. BIOL. 192:177-208(1986).
 CC -!- SIMILARITY: TO H0LF1, H0LF1, H0LF5, H0LF6, AND H0LF7.
 DR EMBL; X04630; HEHCMVU.
 DR PIR; C26079; Q02EC3.
 KW HYPOTHETICAL PROTEIN.
 FT CARBOHYD 76 76 POTENTIAL.
 FT CARBOHYD 113 118 POTENTIAL.
 FT CARBOHYD 223 223 POTENTIAL.
 SQ SEQUENCE 846 AA; 91047 MW; 3448605 CN;

Initial Score = 9 Optimized Score = 69 Significance = 4.34
 Residue Identity = 22% Matches = 86 Mismatches = 248
 Gaps = 46 Conservative Substitutions = 0

X	10	20	30	40	50		
MQTRRVVLKSAAAGITLLGGLAGCATWLDRSA-DAIG-----SIRARPITISEAGFTLTHED							
ASAPHPASLLTAVRRHLNQRLCCGVLAALGAVLPARWLGCAAGPATGTAAGTTSPPAASGTETEAAGGDAPCA	330	X 340	350	360	370	380	390
60	70	80	90	100	110	120	
ICG--SSAGFLRAWPEFFGSRK---LA-EKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVH							
IAGAVGSAVPVPPQPYGAAGGGAICSVNADAHAVVGADAAAAAAAPTVMVGSTAMAGPAAS--GTVPRAMLVV	400	410	420	430	440	450	460
130	140	150	160	170	180		
IVAATGLWF-DPPLSMRLRYVEELTQFFLREIQYQGIEDTG-IRAGIIKVATTGKATPFQE-----LVLKA							
LLDELGAVFGYCPLDGHVYPLAAELSHFLRAGVLGALALGRESAPAAEAARRLLPELDREQWERPRWDALHL	470	480	490	500	510	520	530
190	200	210	220	230	240	250	
AARASLATGVPVTHTAASQRDGE-RGRFFFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHS							
HPRAALWAREP---HGQWEFMFREQRGDPINDPLAFRLSDARTLGLDLTTVMTERQSQLPEKYIGFYQIRKP	540	550	560	570	580	590	600
260	270	280	290	300	310		
AIGLEDNASASPLLGIRSWQT---RALLIKALIDQGYMKQILVSNWLFQFSSYV-----TNIMDVMDRVNP							
PWLME-----QFPPPSRQTKPDAATMPPPLSAQASVSYALRYDDESWRPLSTVDDHKAWLDDLDDESHWVLG	610	620	630	640	650	660	
320	330	340	X				
D-GMAFIPLR-VIPFYERRASHRKRCQASL							
DSRPDDIKQRRLLKATQRRGAEIDRPMPVVPEECYDQRFT	670	680	690	700			

4. LOW344-FIG1. PEP

CYAA\$BORPE CALMODULIN-SENSITIVE ADENYLYL CYCLASE PRECURSOR (

ID CYAA\$BORPE STANDARD: FRT; 1706 AA.
 AC P15318;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
 DE CALMODULIN-SENSITIVE ADENYLYL CYCLASE PRECURSOR (EC 4.6.1.1)
 DE (ATP PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE) (CYCLOLYSIN) (CONTAINS:
 DE HEMOLYSIN) (GENE NAME: CYA).
 OS BORDETELLA PERTUSSIS.
 OC PROKARYOTA; BACTERIA; GRAM-NEGATIVE AEROBIC RODS AND COCCI; UNCERTAIN.
 RN [1] (STRAIN 18323, SEQUENCE FROM N. A.)
 RA GLASER P., LADANT D., SEZER O., PICHOT F., ULLMANN A., DANCHIN A.;
 RL MOL. MICROBIOL. 2:19-30(1988).
 CC -!- FUNCTION: THIS ADENYLYL CYCLASE BELONGS TO A SPECIAL CLASS OF
 CC BACTERIAL TOXIN. IT ACTS ON MAMMALIAN CELLS BY ELEVATING CAMP-
 CC CONCENTRATION AND THUS DISRUPTS NORMAL CELL FUNCTION.
 CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: RELEASED EXTRACELLULARLY IN A PROCESSED
 CC FORM.
 CC -!- DISEASE: WHOOPING COUGH.
 DR EMBL: Y00545; BPCYA.
 DR PIR: S00893; S00893.

DR PROSITE; PS00330; HEMOLYSIN_CALCIUM.
 KW LYASE; CAMP SYNTHESIS; HEMOLYSIS; TOXIN; VIRULENCE; WHOOPING COUGH;
 KW CALCIUM-BINDING; REPEAT.
 FT CHAIN 1 1706 CALMODULIN-SENSITIVE ADENYLYLATE
 CYCLASE PRECURSOR.
 FT CHAIN 1 ? CALMODULIN-SENSITIVE ADENYLYLATE
 CYCLASE.
 FT CHAIN 1 1300 HEMOLYSIN, BY SIMILARITY TO E. COLI
 HEMOLYSINS (HYLA).
 FT DOMAIN 1 360 A, CALMODULIN-SENSITIVE CATALYTIC
 CENTRE.
 FT DOMAIN 361 912 B, ALA/GLY RICH.
 FT DOMAIN 913 1656 C.
 FT DOMAIN 1657 1706 D, ASP/GLY RICH.
 SQ SEQUENCE 1706 AA; 177505 MW; 1.25518E+07 CN;

Initial Score = 10 Optimized Score = 69 Significance = 4.34
 Residue Identity = 23% Matches = 93 Mismatches = 206
 Gaps = 97 Conservative Substitutions = 0

	10	20	30	40		
	X					
	M@TRRVVLKSAAAGTLLGGLAGCATWLDRSAQAIGSIR				ARPITI	---SEA
	:	:	:	:	:	:
	LMTQFGRAGSTNTPQEAAASLSAAVFGL-GEASSAVAETVSGFFRGSSRWAGGFVGAGGAMALGGGIAAAVGA					
490	500	510	520	530	540	550
50	60	70	80	90	100	110
GFTLTHEDICG--SSAGFLRAWPEFFG----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAE						
:	:	:	:	:	:	:
GMSLTDDAPAGQKAAAGAEIALQLTGGLVELASSIALA					LAAARGVTSGLQVAGASAG	
560	570	580	590	600		
120	130	140	150		160	170
VSRAADVHIVAATGLWFDPPLSMRLRYVEELTQFFLREIQYGLIE					DTGIRAGI	IKVATT
:	:	:	:	:	:	:
---AAAGALAAALSPMEIYGLVQQSHYADQLDKLAQESSAYGYEGDALLAQLYRDKTAEGAVAGVSAVLST						
610	620	630	640	650	660	670
180	190	200	210	220	230	
GKATPFQELVLKAAARASLATGVPVTHTAASQRDG----ERGRPPFLSPKLEPSRVCIGHSDDTDDLSYL						
:	:	:	:	:	:	:
VGA----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGVQQPIIEKL----ANDYARKIDELEGGP						
680	690	700	710	720		730
240	250	260	270	280	290	300
TA-----LLRGYLIGLDHIPHSAGILEDNASASPILLGIRSW@TRALLIKALIDQGYMK@ILVSNDWLFGF						
:	:	:	:	:	:	:
QAYFEKNLQARHEQLANSQGLRKMLADLQAGWNASSVIG--V@TTEISKSAL----ELAAITGNADNL--K						
740	750	760	770	780		790
310	320	330	340	X		
SSYVTNIMDVMDRVNPDGMAFIP--LRVIPFYERRASHRKRCQASL						
:	:	:	:	:		
SVDV----FVDRFVQGERVAGI@PVVLDVAAGGIDIAS-RKGERPALTFTITPLAAGP						
800	810	820	830	840		

5. LOW344-FIG1. PEP

CAP1\$MESCR PHOSPHOENOLPYRUVATE CARBOXYLASE 1 (EC 4.1.1.31) (G

ID	CAP1\$MESCR	STANDARD:	PRT:	366 AA.
AC	P10490;			
DT	01-JUL-1989	(REL. 11, CREATED)		
DT	01-JUL-1989	(REL. 11, LAST SEQUENCE UPDATE)		
DT	01-APR-1990	(REL. 14, LAST ANNOTATION UPDATE)		
DE	PHOSPHOENOLPYRUVATE CARBOXYLASE 1 (EC 4.1.1.31) (GENE NAME: PPCA).			

OS COMMON ICE PLANT (MESEMBRYANTHEMUM CRYSTALLINUM).
 OC EUKARYOTA; PLANTAE; SPERMATOPHYTA; ANGIOSPERMAE.
 RN [1] (SEQUENCE FROM N. A.)
 RA RICKERS J., CUSHMAN J., MICHALOWSKI C., SCHMITT J., BOHNERT H. J.;
 RL MOL. GENET. 215:447-454 (1989).
 RN [2] (SEQUENCE FROM N. A.)
 RA CUSHMAN J. C., BOHNERT H. J.;
 RL NUCLEIC ACIDS RES. 17:6745-6745 (1989).
 CC -!- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.
 CC -!- CATALYTIC ACTIVITY: ORTHOPHOSPHATE + OXALOACETATE = H(2)O +
 CC PHOSPHOENOLPYRUVATE + CO(2).
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 DR EMBL; X13660; MCPPCR.
 DR EMBL; X14587; MCPPCA.
 KW LYASE; CARBON DIOXIDE FIXATION; ALLOSTERIC ENZYME;
 KW TRICARBOXYLIC ACID CYCLE.
 SQ SEQUENCE 366 AA; 110653 MW; 4690045 CN;

Initial Score = 8 Optimized Score = 69 Significance = 4.34
 Residue Identity = 23% Matches = 94 Mismatches = 212
 Gaps = 90 Conservative Substitutions = 0

X	10	20	30	40	50	
MQTRRVVLKSAAAGTLLGGLAGCATWLRSAA--IGSIR---ARPITISEAGFTLTHEDI						
	;	;	; ; ; ;	; ; ; ;	; ; ; ;	
SVRRSLLQKHGRIRDCLQLYAKDITPDKQELDEALQREIQAQFRDDEIRRTQPTPQDEMRAAGMSYFHETI						
180	180	200	210	220	230	
60	70	80		90	100	110
CGSSAGFLRAWPEFFGSRKALAEKAVRGLRAR-----AAGVRTIVDVSTFDIGRDVSLLAEVSR						
	; ; ; ; ; ;			; ; ; ; ; ;		; ; ; ; ; ;
WNGVPKFLR-----RLDTALK-NIGITERVPYNAPLIQFSSWMGGDRDGNPRVTPEVTRDVCLLA-RMM						
250	260	270	280	290	300	310
120	130	140	150	160	170	
AADVHIVAATGLWFDPPPLSM-----RLR-YVEELTQFFLREIQYGIEDTGIRAGIIVKATTGKATPFQE--L						
	; ; ; ; ; ;	; ; ; ; ; ;		; ; ; ; ; ;		; ; ; ; ; ;
AANMYFSQIDELMF--ELSMWRCTDELRERAEELHKYSKRDSDKHYIE-----FWKQIIPSSEPYR						
	320	330	340	350		360
180	190	200	210	220	230	
VLKA-----AARASLATGV---PV-TTHTAASQRDGERGRPPFLSP-KLEPSRVCIGHSDDTDDLS						
	; ; ; ; ; ;	; ; ; ; ; ;		; ; ; ; ; ;		; ; ; ; ; ;
VILADVRDKLYYTRERSRQLLASEVSEIPVEATFTEIDQ-----FLEPLECYRSLCACGDRPVADGS						
370	380	390	400	410	420	430
240	250	260	270	280	290	300
YL-----TALLRGYLIGLDHIPHSAGLEDNASASPLLGIRSWQTRALLIKALIDQGYMKQILVSNDWLFGF						
	; ; ; ; ; ;		; ; ; ; ; ;		; ; ; ; ; ;	
LLDFMRQVATFGLCLVKLDIQRQESERHTDVMDAITTHLGIGS--YRDWTEEKRQD--WLLSELRGKRPLFGP						
440	450	460	470	480	490	
310	320	330	340	X		
SSYVT-NIMDVMDRVNPDGMAFIPLRVIPFY-ERRASHRKRCQASL						
	; ; ; ; ; ;					
DLPRRTDEIADVLDTIN--VIAELPSDSFGAYVISMAAPSVDLAVELLQRECKVKK						
500	510	520	530	540	X	550

6. LOW344-FIG1. PEP
 MIS\$HUMAN MULLERIAN INHIBITING FACTOR PRECURSOR (MIS).

ID MIS\$HUMAN STANDARD: PRT: 560 AA.
 AC P03971;
 DT 23-OCT-1986 (REL. 02, CREATED)

320	330	340	X		
MDRVNPDG	AFIPLRVIPFY	ER-RASHRKRCBASL			
NHVVLLLKM Q ARGAALARPPCCVPTAYAGKLLISLSEERI SAHH VPNM VATECGCR					
510	520	530	540	550	X 560

7. LOW344-FIG1. PEP

ATXA\$LEIDO PROBABLE E1-E2 TYPE CATION ATPASE 1A (EC 3.6.1.-).

ID ATXA\$LEIDO STANDARD: PRT: 974 AA.
 AC P11718;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)
 DE PROBABLE E1-E2 TYPE CATION ATPASE 1A (EC 3.6.1.-).
 OS LEISHMANIA DONOVANI.
 OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA.
 RN [1] (SEQUENCE FROM N. A.)
 RA MEADE J. C., SHAW J., LEMASTER S., GALLAGHER G., STRINGER J. R.;
 RL MOL. CELL. BIOL. 7:3937-3946(1987).
 CC -!- CATALYTIC ACTIVITY: ATP + H₂O = ADP + ORTHOPHOSPHATE.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).
 CC -!- SIMILARITY: THE TWO L. DONOVANI CATION-TRANSPORTING ATPASE GENES
 CC ARE 98% HOMOLOGOUS.
 CC -!- CAUTION: IN POSITION 351 THE N. A. SEQUENCE PREDICTS ARG, THE
 CC PROTEIN TRANSLATION SHOWN IN THREE PLACES IN THE PAPER GIVES
 CC LYS, WHICH IS CONSERVED IN ALL KNOWN E1-E2 ATPASES. WE HAVE
 CC USED LYS AFTER CONFIRMATION FROM THE AUTHORS.
 DR EMBL; M17889; LDCATP1.
 DR PROSITE; PS00154; ATPASE_E1_E2.
 KW HYDROLASE; ATP HYDROLYSIS; TRANSMEMBRANE; PHOSPHORYLATION;
 KW MAGNESIUM; ATP-BINDING.
 FT TRANSMEM 93 112 PUTATIVE.
 FT TRANSMEM 118 137 PUTATIVE.
 FT TRANSMEM 265 286 PUTATIVE.
 FT TRANSMEM 295 321 PUTATIVE.
 FT TRANSMEM 631 651 PUTATIVE.
 FT TRANSMEM 662 684 PUTATIVE.
 FT TRANSMEM 698 712 PUTATIVE.
 FT TRANSMEM 738 761 PUTATIVE.
 FT TRANSMEM 813 840 PUTATIVE.
 FT TRANSMEM 869 887 PUTATIVE.
 FT MOD_RES 351 351 PHOSPHORYLATION.
 SQ SEQUENCE 974 AA; 107448 MW; 5115862 CN;

Initial Score = 18 Optimized Score = 68 Significance = 3.95
 Residue Identity = 24% Matches = 93 Mismatches = 210
 Gaps = 82 Conservative Substitutions = 0

	X	10	20	30	40	50	
	MQTRR	-----VVLKSAAAGTLLGGLAGCATWLDRSAQAIGSIRARPITISEAGFTLTHE					
FLDPPRPDTKDT	IRR	SKEYGVDVKM	ITGDHILLIAKEMC	-RM	LDLDPNILTAD	KLPOIKDANDLPEDLGEKYG	
500	510	520	530	540	550	560	
	60	70	80	90	100	110	120
DICGSSAGFLRAWPEFFGSRKALAIKEAVRGLRAR	---	AAGVRTIVDV	VSTFDIGRDV	SLLAEVSRAADVHIVA			
DMML	SVGGFAQVFPE	-----HKF	MIVETLRQ	RGYTCAMTGDGV	NDA	PALKR	ADV
570	580	590	600	610	620	630	
	130	140	150	160	170		
ATC	LHERRDPLGMEV	PMV	EL	TOEEL	PISTOV	VIERT	

8. LOW344-FIG1. PEP

POLG\$WNV GENOME POLYPROTEIN (CAPSID PROTEIN C; ENVELOPE GLY

ID POLG\$WNV STANDARD; PRT; 3430 AA.
 AC P06935;
 DT 01-JAN-1988 (REL. 06, CREATED)
 DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
 DT 01-MAR-1989 (REL. 10, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN (CAPSID PROTEIN C; ENVELOPE GLYCOPROTEIN M; MAJOR
 DE ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS3, NS4A,
 DE NS4B AND NS5).
 OS WEST NILE VIRUS.
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; FLAVIVIRIDAE.
 RN [1] (SEQUENCE FROM N. A.)
 RA CASTLE E., LEIDNER U., NOWAK T., WENGLER G., WENGLER G.;
 RL VIROLOGY 149:10-26(1986).
 RN [2] (SEQUENCE OF 1-291 FROM N. A.)
 RA CASTLE E., NOWAK T., LEIDNER U., WENGLER G., WENGLER G.;
 RL VIROLOGY 145:227-236(1985).
 RN [3] (SEQUENCE OF 255-854 FROM N. A.)
 RA WENGLER G., CASTLE E., LEIDNER U., NOWAK T., WENGLER G.;
 RL VIROLOGY 147:264-274(1985).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND mRNA.
 DR PIR; A25256; GNWVW.
 DR EMBL; M10103; FLWNVSP.
 KW POLYPROTEIN; COAT PROTEIN; ENVELOPE PROTEIN; NONSTRUCTURAL PROTEIN;
 KW TRANSMEMBRANE; GLYCOPROTEIN.
 FT CHAIN 1 105 CAPSID PROTEIN C.
 FT PROPEP 106 215
 FT CHAIN 216 290 ENVELOPE GLYCOPROTEIN M.
 FT CHAIN 291 787 MAJOR ENVELOPE PROTEIN E.
 FT CHAIN 788 1187 NONSTRUCTURAL PROTEIN NS1.
 FT CHAIN 1188 1354 NONSTRUCTURAL PROTEIN NS2A.
 FT CHAIN 1355 1484 NONSTRUCTURAL PROTEIN NS2B.
 FT CHAIN 1485 2109 NONSTRUCTURAL PROTEIN NS3.
 FT CHAIN 2110 2394 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 2395 2579 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2580 3430 NONSTRUCTURAL PROTEIN NS5.
 FT CARBOHYD 138 138 POTENTIAL.
 FT CARBOHYD 917 917 POTENTIAL.

FT CARBOHYD 962 562 POTENTIAL.
 FT CARBOHYD 994 934 POTENTIAL.
 FT CARBOHYD 1283 1289 POTENTIAL.
 FT CARBOHYD 1659 1659 POTENTIAL.
 FT CARBOHYD 2336 2336 POTENTIAL.
 FT CARBOHYD 2489 2489 POTENTIAL.
 FT CARBOHYD 2573 2573 POTENTIAL.
 FT CARBOHYD 2739 2739 POTENTIAL.
 FT CARBOHYD 2759 2759 POTENTIAL.
 FT CARBOHYD 2864 2864 POTENTIAL.
 FT CARBOHYD 2902 2902 POTENTIAL.
 SQ SEQUENCE 3430 AA: 379840 MW: 2.098737E+07 CN:

 Initial Score = 10 Optimized Score = 68 Significance = 3.95
 Residue Identity = 21% Matches = 87 Mismatches = 231
 Gaps = 78 Conservative Substitutions = 0

 X 10 20 30 40 50
 MQTRRVVLKSAAGTLLGGLAGCAT-----WLDRSAQAIGSIRARPITISEAGFTLTHE-
 | :
 VESHGKIGATQAGRFSITPSAPSYTTLKLGEGYGEVTVDCEPRSGIDTSAYYVMS-----VGEKSFLVHREW
 440 450 460 470 480 490 500

 60 70 80 90 100
 ----DIGSSAG-----FLRAWPE---FFGSRKAL--AEKAVRGLRARAAGV---RTIVDVSTFDIGR
 | :
 FMDLNLPWSSAGSTTWRNRETLMEFEPEHATKQSVVALGSQEGALH8ALAGAIPVEFSSNTVKLTSGHLKCR
 510 520 530 540 550 560 570

 110 120 130 140 150 160 170
 DVSLLAEVSRRAADVHIVAATGLWFDPPPLSMRLRYVEELTQFFLREIQYGIEDTGIRAGIIKVATTGKATPFO
 | :
 VKMEKLQLKGTTYGVCSSKAFFKARTPADTGHGTVVLEL-----QYTGTDGPCKVPISSVASLNDLTPVG
 580 590 600 610 620 630

 180 190 200 210 220 230
 ELVLKAAARASLATGVP-VTTHTAASQRDGERGRPPFLS-----PKLEPSRVCIGHSDDT-----
 | :
 RLV-----TVNPFPVSVATANSKVLIELEPPFGDSYIVVGRGEQQINHHWHKSGSSIGKAFTTTLRGA
 640 650 660 670 680 690 700

 240 250 260 270 280 290 300
 DDLSYL--TALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWQTRALLIKALIDQGYMKQILVSNDWLF
 | :
 QRLAALGDTAWDFGSVGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLL-GALLLWMGINARDRSIAMTFL
 710 720 730 740 750 760 770

 310 320 330 340 X
 FSSYVTNIMDVMDRVNPD-GMAFIPLRVIPFYERRASHRKRCQASL
 | :
 AVGGVLLFLSV--NVHADTGCAIDIGRQELRCGSGGVFIHNDVEAWMDRYKFYPETP
 780 790 800 810 X 820

9. LOW344-FIG1. PEP

ATXB\$LEIDO PROBABLE E1-E2 TYPE CATION ATPASE 1B (EC 3.6.1.-).

ID ATXB\$LEIDO STANDARD: PRT: 974 AA.
 AC P12522;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)
 DE PROBABLE E1-E2 TYPE CATION ATPASE 1B (EC 3.6.1.-).
 OS LEISHMANIA DONOVANI.
 OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA.
 RN [1] (SEQUENCE FROM N. A.)

RA MEADE J. C., HUDSON K. M., STINGER S. L., STRINGER J. R. ;
 RL MOL. BIOCHEM. PARASITOLOGY 33:81-92(1989).
 CC -!- CATALYTIC ACTIVITY: ATP + H₂O = ADP + ORTHOPHOSPHATE.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).
 CC -!- SIMILARITY: THE TWO L. DONOVANI CATION-TRANSPORTING ATPASE GENES
 CC ARE 98% HOMOLOGOUS.
 DR EMBL: J04004; LDCATP2.
 DR PROSITE: PS00154; ATPASE_E1_E2.
 KW HYDROLASE; ATP HYDROLYSIS; TRANSMEMBRANE; PHOSPHORYLATION;
 KW MAGNESIUM; ATP-BINDING.
 FT TRANSMEM 93 112 PUTATIVE.
 FT TRANSMEM 118 137 PUTATIVE.
 FT TRANSMEM 265 286 PUTATIVE.
 FT TRANSMEM 295 321 PUTATIVE.
 FT TRANSMEM 631 651 PUTATIVE.
 FT TRANSMEM 662 684 PUTATIVE.
 FT TRANSMEM 698 712 PUTATIVE.
 FT TRANSMEM 738 761 PUTATIVE.
 FT TRANSMEM 813 840 PUTATIVE.
 FT TRANSMEM 869 887 PUTATIVE.
 FT MOD_RES 351 351 PHOSPHORYLATION.
 SQ SEQUENCE 974 AA; 107304 MW; 5132373 CN;

Initial Score = 18 Optimized Score = 68 Significance = 3.95
 Residue Identity = 24% Matches = 93 Mismatches = 210
 Gaps = 82 Conservative Substitutions = 0

	X	10	20	30	40	50			
	MQTRR	-----VVLKSAAGTLLGGLAGCATWLDRSAQAIGSIRARPITISEAGFTLTHE	-----						
500	FLDPPRPDTKDTIRRKEYGVVKMITYCDHLLIAKEMC-RMLDLDPNILTADKLPQIKDANDLPEDLGEKYG	510	520	530	540	550	560		
	60	70	80	90	100	110	120		
	DICGSSAGFLRAWPEFFGSRKALAEKAVRGLRAR	-----AAGVRTIVDVSTFDIGRDVSVLLAEVSRAADVHIVA	-----						
570	DMMLSVGGFAQVFPE	-----HKFMIVETLRQRGYTCAMTGDGVNDAPALKRADV	--GIAVHGATDAARAA	580	590	600	610	620	630
	130	140	150	160	170				
	ATGLWFDPPLSMRLRYVE	-----ELTQFFLREIQYGIEDT	---GIRAGIIVKVATTGKA	-----TPFQ	---E				
640	ADMVLTEPGLS	-----VVVEAMLVSREVFRMLSLTYRISATLQLVCFFFIA	CFSLTPKAYGSVDPNFQFFHL	650	660	670	680	690	700
180	190	200	210	220	230	240			
	LVLKAAARASLATGVPVT	-----THTAASGRDGERGRPPFLSPKLEPSRVCIGH	SDDTDDLSYL	TALLRGY	LIG				
710	PVLMFMLITLLNDGCLMTIGYDHVIP	-----ERPQKWNL	-PVVFVS	-----	-----	ASILAAVACGSSL	M		
	720	730	740			750			
	250	260	270	280	290	300	310		
	LDHIPHSAGILEDNASASPLLGIRSWQTR	-----ALLIKALIDQGYMKQILVSN	WDLFGFSS	--YVTNIMDVM					
760	L-----LWIGLE	-----GYSSQYYENSWFHRLGLAQLPQGKLV	TMMYLK	-ISIS	-DFLTLFSSRTGGHFFFYVP				
	770	780	790			800	810		
	320	330	340	X					
	DRVNPDGMAFIPLRVIPFYERRAS	--HRKRCQASL							
820	PSPILFCGAIISLLV	-----STMAASFWHKS	RPDNVLTEGLAWCQTN						
	830	840	850			860			

ANALYSIS OF THE SYNTHETIC ALPHA CHAIN, MITOCHONDRIAL (EC 3.6.1.1).

ID ATP00ENBI STANDARD: PRT: 511 AA.
AC P05492;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)
DE ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL (EC 3.6.1.34).
OS OENOTHERA BIENNIS.
OG MITOCHONDRION.
OC EUKARYOTA; PLANTA; SPERMATOPHYTA; ANGIOSPERMAE.
RN [1] (SEQUENCE FROM N. A.)
RA SCHUSTER W., BRENNICKE A.;
RL MOL. GEN. GENET. 204:23-35(1986).
CC -!- FUNCTION: THIS IS ONE OF THE 5 CHAINS OF THE ENZYMATIC COMPONENT
CC (COUPLING FACTOR CF(1)) OF THE MITOCHONDRIAL ATPASE COMPLEX.
DR EMBL; X04023; MIOBATPA.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA.
KW ATP SYNTHESIS; CF(1) COUPLING FACTOR; HYDROGEN ION TRANSPORT;
KW HYDROLASE; ATP-BINDING; MITOCHONDRION.
FT NP_BIND 171 178 ATP (BY SIMILARITY).
FT ACT_SITE 373 373 BY HOMOLOGY.
SQ SEQUENCE 511 AA; 55596 MW; 1250759 CN;

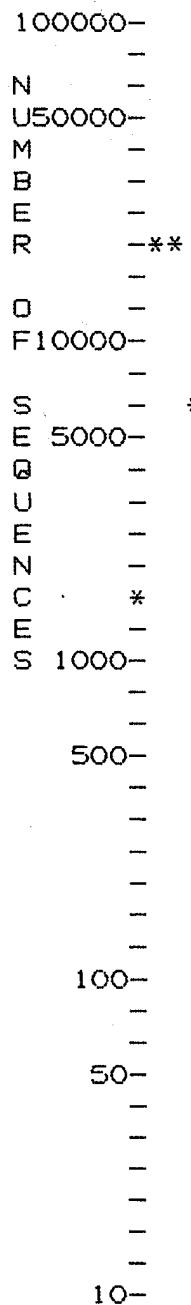
Initial Score = 9 Optimized Score = 68 Significance = 3.95
Residue Identity = 23% Matches = 91 Mismatches = 223
Gaps = 67 Conservative Substitutions = 0

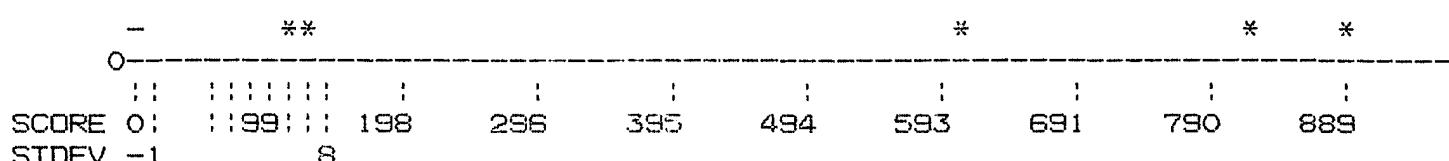
X	10	20	30	40	50	60	
M	QTRRVVLKSAAGTLLGGLAGCATWLDRSAQ--AIG-SIR----ARPITISEAGFTLT	EDICGSSAGFL					
	MEFS	SPRAAELT	TLLESRIT	NFYTNFQVDEIGRVISVGDDGIARVYGLNEI	QAGEMVEFASGVKGIAL		
X	10	20	30	40	50	60	
	70	80	90	100	110	120	
RAWPE	----FFGSRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVHI--VAATG---						
NLENENVGI	VVFGSDTAIKE	---GDLVKR	---TGSIVDV	----PAGKSLLGRVVDALGVPIDGRGALGDHE			
70	80	90	100	110	120		
	130	140	150	160	170	180	190
---	LWFDPPPLSMRLRYVEELTQFFL	-REI	QYGIEDTGIRAGI	IKVATTGKATPFQELVLKAAARASLATGVP			
RRRVEVKVPGI	IERKSVHEPMQTGILKAVDSLVP	FIGRGQRELI	IGDRQTGKTAIAIDT	ILNQKQMNSRATSES			
130	140	150	160	170	180	190	
	200	210	220	230	240	.	250
VT	---THTAASQ	-RDGERGRPPFLS	--PKLEPSRVCIGHSDDTDDLSYLT	TALLRGYLIG	---LDHIPHSAI-		
ETLYCVYVAIG	QKRSTVAQVLVDI	SEGNALEYSILVAATASDPAPLQFL	-APYSGCAMGEYFRDNGMHALII				
200	210	220	230	240	250	260	270
	260	270	280	290	300	310	
--	GLEDNASA	--SPLLGIRSWQTRALLIKALIDQGYMKQIL	----VSNDWLFGFSSYVTNIMDVMDRVNPD				
YDDLSKQAVAYRQMSLLRRPPGREA	---	FPGDVFYLHSRLLERA	AKRSQQTGAGS	--LTALPVIETQAGD			
280	290	300	310	320	330		
	320	330	340	X			
GMAFIPLRVIPFYERRASHRKRCBASL							
VSAYIPTNVI	-----SITDGQX	CLETELFYRGIR					
340	350	X	360				

Results file low344-fig1.res made by maryh on Wed 17 Apr 91 12:41:07-PDT.

Query sequence being compared: LOW344-FIG1. SEQ
Number of sequences searched: 43406
Number of scores above cutoff: 156

Results of the initial comparison of LOW344-FIG1. SEQ with:
Data bank : GenBank 65.0, all entries
Data bank : UEMBL 24_65, all entries





PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	79		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	31	31	12.75
Times:	CPU		Total Elapsed
	00:54:30.02		02:41:00.00
Number of residues:	54775335		
Number of sequences searched:	43406		
Number of scores above cutoff:	156		

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame	
***** 67 standard deviations above mean *****							
1. PSEPTE	Plasmid pCMS1 (from <i>P. diminuta</i>	1322	889	1304	67.29	0	
***** 61 standard deviations above mean *****							
2. M22863	Figure 1. Nucleotide sequence	1326	809	1309	61.02	0	
***** 44 standard deviations above mean *****							
3. FVBOPD	<i>Flavobacterium</i> sp. parathion h	1693	599	1281	44.55	0	
***** 8 standard deviations above mean *****							
4. X15898	<i>Eimeria tenella</i> mRNA for sporo	957	136	360	8.23	0	
***** 6 standard deviations above mean *****							
5. X14805	Mouse mRNA encoding DNA (cytos	4973	112	569	6.35	0	
6. CELPOLII	<i>C. elegans</i> RNA polymerase II 1a	12993	111	526	6.27	0	
7. HSHEPSH	Human hepatoma mRNA for serine	2363	108	562	6.04	0	
8. HUMHPSNA	Human hepatitis mRNA, complete cd	1783	108	456	6.04	0	
***** 5 standard deviations above mean *****							
9. BLYAMY2	Barley (<i>H. vulgare</i>) alpha-amylase	1588	107	590	5.96	0	

10.	BLYAMYAA	Barley alpha-amylase type A is	1588	107	592	5. 96	0
11.	BOVGABARB	Bovine mRNA for gamma-aminobut	3010	106	574	5. 88	0
12.	BOVIGCAB	Bovine Ig germline gamma-2-cha	1979	104	479	5. 73	0
13.	BTIGG2HC	Bovine Ig germline heavy chain	1979	104	479	5. 73	0
14.	RRATP2	Rhodospirillum rubrum gene clu	4240	104	184	5. 73	0
15.	PDEMDH	P. denitrificans methanol dehyd	2314	103	425	5. 65	0
16.	MUSGT2A	M. musculus glucose transporter	2521	102	232	5. 57	0
17.	SMASFUABC	S. marcescens periplasmic-bind	4583	101	583	5. 49	0
18.	RABIGHAB	Rabbit Ig mu chain secreted fo	1953	101	489	5. 49	0
19.	HUMIGHBD	Human Ig unproductively rearr	1127	100	384	5. 41	0
20.	MUSRGE3	Mouse 18S, 5. 8S, 28S rRNA gene	3061	98	360	5. 25	0

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Init. Opt.					
		Length	Score	Score	Sig.	Frame	
**** 271 standard deviations above mean ****							
1. M22863	Figure 1. Nucleotide sequence	1326	809	1309	271. 26		0
**** 269 standard deviations above mean ****							
2. PSEPTE	Plasmid pCMS1 (from P. diminuta	1322	889	1304	269. 45		0
**** 261 standard deviations above mean ****							
3. FVBOPD	Flavobacterium sp. parathion h	1693	589	1281	261. 10		0
**** 15 standard deviations above mean ****							
4. TRN21TNPA	Transposon Tn21 tnpA gene for	3176	87	606	15. 98		0
**** 12 standard deviations above mean ****							
5. X17379	Sorghum vulgare mRNA for phosph	3147	80	597	12. 71		0
**** 11 standard deviations above mean ****							
6. NEUTRP1	n. crassa trifunctional tryptop	2750	85	594	11. 62		0
**** 10 standard deviations above mean ****							
7. BLYAMYAA	Barley alpha-amylase type A is	1588	107	592	10. 89		0
8. PDUMER	Plasmid pDU1358 (from S. marces	2153	95	590	10. 17		0
9. BLYAMY2	Barley (H. vulgare) alpha-amyla	1588	107	590	10. 17		0
10. BPECYADE	Bordetella pertussis cyaD gene	2040	82	590	10. 17		0
**** 8 standard deviations above mean ****							
11. MUSNFMG	Mouse NF-M gene for middle-mol	5471	81	586	8. 72		0
**** 7 standard deviations above mean ****							
12. ACFTS140A	Fujinami sarcoma virus tempera	2715	86	584	7. 99		0
13. SVGSII	Streptomyces viridochromogenes	2755	82	584	7. 99		0
14. SMASFUABC	S. marcescens periplasmic-bind	4583	101	583	7. 63		0
15. HS11UL	Herpes simplex virus type 1 (H	108360	94	582	7. 26		0
**** 6 standard deviations above mean ****							
16. ATUPRIREP	A. tumefaciens plasmid pRiA4b r	4638	80	581	6. 90		0
17. HUMASPX	Human nonerythroid alpha-spect	7787	80	581	6. 90		0
18. SERCYSA	S. erythraea rhodanese-like pro	3373	81	581	6. 90		0
19. X51950	E. coli purHD operon for AICAR	3432	94	580	6. 54		0
20. MUSHCK	Mouse hck gene for tyrosine ki	1960	81	580	6. 54		0

1. LOW344-FIG1. SEQ

M22863 Figure 1. Nucleotide sequence of Flavobacterium op

LOCUS M22863 1326 bp ds-DNA UNA 15-JUN-1989
 DEFINITION Figure 1. Nucleotide sequence of Flavobacterium opd gene fragment.
 ACCESSION M22863
 KEYWORDS .
 SOURCE
 ORGANISM Unknown
 Unclassified.
 REFERENCE 1 (bases 1 to 1326)

AUTHORS Harper, L. L., McDaniel, C. S., Miller, C. E. and Wild, J. R.
 TITLE Dissimilar Plasmids Isolated from *Pseudomonas diminuta* MG and a
Flavobacterium sp. (ATCC 27551) Contain Identical *opd* Genes
 JOURNAL Appl. Environ. Microbiol. 54, 2586-2589 (1988)
 STANDARD unannotated staff_entry
 BASE COUNT 279 a 363 c 392 g 286 t
 ORIGIN

Initial Score = 809 Optimized Score = 1309 Significance = 271.26
 Residue Identity = 98% Matches = 1316 Mismatches = 6
 Gaps = 7 Conservative Substitutions = 0

X 10 20 30 40 50 60 70
 CTGCAGCCTGACTCGGCACCAAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

CTGCAGCCTGACTCGGCACCAAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
 X 10 20 30 40 50 60 70

80 90 100 ↓ 110 ↓ 120 130 140
 GAAGGGTTGTGCTCAAGTCTGCGGUCGCGAGAACCTCTGCTGGCGGCCTGGCTGGGTGCGCGACGTGGCTGG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

GAAGGGTTGTGCTCAAGTCTGCGGUCGCGAGAACCTCTGCTGGCGGCCTGGCTGGGTGCGCGACGTGGCTGG
 80 90 100 110 120 130 140

150 160 ↓ 170 180 190 200 210
 ATCGATCGGCACAGGCATCGGATCAATACTCGTGCCTATCACAACTCTGAAGCGGGTTTCACACTGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

ATCGATCGGCACAGGCATCGGATCAATACTCGTGCCTATCACAACTCTGAAGCGGGTTTCACACTGA
 150 160 170 180 190 200 210

220 230 240 250 260 270 280
 CTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTTGCCTGGCCAGAGTTCTCGGTAGCCGCAAAG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

CTCACGAGGACATCTCGGCAGCTCGGCAGGATTCTTGCCTGGCCAGAGTTCTCGGTAG-CGCAAAG
 220 230 240 250 260 270 280

290 ↓ 300 310 320 330 340 350 360
 CTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGGCGTGCAGATTGTCGATGTGCGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

CTCTAGCGCAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGGCGTGCAGATTGTCGATGTGCGA
 290 300 310 320 330 340 350

370 380 390 400 410 420 430
 CTTTCGATATCGGTGCGACGTCAGTTATTGGCCGAGGTTTCGCGGGCTGCCGACGTTCATATCGTGGCGG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

CTTTCGATATCGGTGCGACGTCAGTTATTGGCCGAGGTTTCGCGGGCTGCCGACGTTCATATC-TGGCGG
 360 370 380 390 400 410 420

440 450 460 470 480 490 500
 CGACCGGCTTGTGGTTCGACCCGCCACTTCGATGCGATTGAGGTATGTAGAGGAACCTCACAC-AGTTCTTC
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

CGACCGGCTTGTGGTTCGACCCGCCACTTCGATGCGATTGAGGTATGTAGAGGAACCTCACACTAGTTCTTC
 430 440 450 460 470 480 490 500

510 520 530 540 550 560 570
 CTGC-GTGAGATTCAATATGCGATCGAAG-ACACCCGGAAATTAGGGCGGGCATTATCAAGGTGCGACCAACAG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

CTGCGGTGAGATTCAATATGCGATCGAAGTACACCCGGAAATTAGGGCGGGCATTATCAAGGTGCGACCAACAG
 510 520 530 540 550 560 570

580 590 600 610 620 630 640
 GCAAGGGCGACCCCGCTTCAGGAGTTAGTGTAAAGGCGGCCGCCGGGCCAGCTTGGCCACCGGTGTTCCGG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::

GCAAGGGCGACCCCGCTTCAGGAGTTAGTGTAAAGGCGGCCGCCGGGCCAGCTTGGCCACCGGTGTTCCGG
 580 590 600 610 620 630 640

650 660 670 680 690 700 710
 TAACCACTCACACGGCAGCAAGTCAGCAGCGATGGTGGCGAGGCAGGCCGCCCCATTTTGAGTCGAAGCTTG
 650 660 670 680 690 700 710
 720 730 740 750 760 770 780
 AGCCCTCACGGGTTTGTATTGGTCACACCGATGATACTGACGATTGAGCTATCTCACCGCCCTGCTGCGCG
 720 730 740 750 760 770 780
 790 800 810 820 830 840 850 860
 GATACCTCATCGGTCTAGACCACATCCCGCACAGTGCAGTTGGCTAGAAAGATAATGCGAGTGCATCACCGC
 790 800 810 820 830 840 850 860
 870 880 890 900 910 920 930
 TCCTGGGCATCCGTTGGCAAAACACGGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGCTACATGAAAC
 870 880 890 900 910 920 930
 940 950 960 970 980 990 1000
 AAATCCTCGTTGAAATGACTGGCTGGGGTTTCGAGCTATGTCACCAACATCATGGACGTGATGGATC
 AAATCCTCGTTGAAATGACTGGCTGGGGTTTCGAGCTATGTCACCAACATCATGGACGTGATGGATC
 940 950 960 970 980 990 1000
 1010 1020 1030 1040 1050 1060 1070
 GCGTGAACCCCGACGGGATGGCCTTCATTCCACTGAGAGTGAATCCCATTCTACGAGAGAAGGGCGTCCCACA
 1010 1020 1030 1040 1050 1060 1070
 1080 1090 1100 1110 1120 1130 1140
 GGAAACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGGTTCTGTGTCACCGACTTGCCGTGCATGACGCC
 1080 1090 1100 1110 1120 1130 1140
 1150 1160 1170 1180 1190 1200 1210 1220
 ATCTGGATCCTTCCACGCAGCGGGCCACTATTCCCCGTCAAGATAACCGAACGATGAAGTCGCGCATCGATCGA
 1150 1160 1170 1180 1190 1200 1210 1220
 1230 1240 1250 1260 1270 1280 1290
 TAGGCATCTTCAATGTGATCAGGGCTGCCACCTCCAAAGGCCGCGGTGGCCACCCCTGTCGATAGTCTTGAGGG
 1230 1240 1250 1260 1270 1280 1290
 1300 1310 1320 X
 CGGTAGCGACGACCGTGCTTTCTGAACTGCAG
 1300 1310 1320 X

2. LOW344-FIG1. SEQ

PSEPTE Plasmid pCMS1 (from *P. diminuta*) phosphodiesterase

LOCUS PSEPTE 1322 bp ds-DNA BCT 15-MAR-1989
 DEFINITION Plasmid pCMS1 (from *P. diminuta*) phosphodiesterase (opd) gene,

ACCESSION M20392
 KEYWORDS phosphotriesterase.
 SOURCE Plasmid pCMS1 (from *Pseudomonas diminuta*) DNA.
 ORGANISM Plasmid pCMS1
 PROKARYOTA; Bacteria; Plasmid pCMS1.
 REFERENCE 1 (bases 1 to 1322)
 AUTHORS McDaniel, C. S., Harper, L. L. and Wild, J. R.
 TITLE Cloning and sequencing of a plasmid-borne gene (opd) encoding a phosphotriesterase
 JOURNAL J. Bacteriol. 170, 2306-2311 (1988)
 STANDARD simple staff_review
 FEATURES Location/Qualifiers
 CDS 63..1040
 /note="phosphotriesterase protein"
 BASE COUNT 278 a 367 c 392 g 285 t
 ORIGIN 5 bp upstream from PstI site.

 Initial Score = 889 Optimized Score = 1304 Significance = 269.45
 Residue Identity = 98% Matches = 1313 Mismatches = 6
 Gaps = 11 Conservative Substitutions = 0

 X 10 20 30 40 50 60 70
 CTGCAGCCTGACTCGGCACCAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 CTGCAGCCTGACTCGGCACCAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
 X 10 20 30 40 50 60 70

 80 90 100 110 120 130 140
 GAAGGGTTGTGCTCAAGTCTGCGGCCGAGGAACCTCTGCTGGCGGCCTGGCTGGGTGCGCGACGTGGCTGG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 GAAGGGTTGTGCTCAAGTCTGCGGCCGAGGAACCTCTGCTGGCGGCCTGGCTGGGTGCGCGACGTGGCTGG
 80 90 100 110 120 130 140

 150 160 170 180 190 200 210
 ATCGATCGGCACAGGGGATCGGATCAATACTCGTGCAGCGTCTATCACAATCTCTGAAGCGGGTTTCACACTGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 ATCGATCGGCACAGGGGATCGGATCAATACTCGTGCAGCGTCTATCACAATCTCTGAAGCGGGTTTCACACTGA
 150 160 170 180 190 200 210

 220 230 240 250 260 270 280
 CTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTGCGTGCCTGGCCAGAGTTCTCGGTAGCCGCAAAG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 CTCACGAGGACATCTCGGCAGGATTCTGCGTGCCTGGCCAGAGTTCTCGGTAGCCGCAAAG
 220 230 240 250 260 270 280

 290 300 310 320 330 340 350 360
 CTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGGCGTGCAGATTGTCGATGTGTCGA
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 CTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGGCGTGCAGATTGTCGATGTGTCGA
 290 300 310 320 330 340 350

 370 380 390 400 410 420 430
 CTTTCGATATCGGTGCGACGTCAGTTATTGGCCGAGGTTCGCGGGCTGCCGACGTTCATATCGTGGCGG
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 CTTTCGATATCGGTGCGACGTCAGTTATTGGCCGAGGTTCGCGGGCTGCCGACGTTCATATCGTGGCGG
 360 370 380 390 400 410 420

 440 450 460 470 480 490 500
 CGACCGGCTTGTGGTTCGACCCGCCACTTTCGATGCGATTGAGGTATGTAGAGGAACACAC-AGTTCTTC
 ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: ::::::::::::::::::::: :::::::::::::::::::::
 CGACCGGCTTGTGGTTCGACCCGCCACTTTCGATGCGATTGAGGTATGTAGAGGAACACACTAGTTCTTC
 430 440 450 460 470 480 490 500

 510 520 530 540 550 560 570
 CTGC-GTGAGATTCAATATGGCATCGAAG-ACACCGAAATTAGGGCGGGCATTATCAAGGTGCGGACACAG

510 520 530 540 550 560 570
 580 590 600 610 620 630 640
 GCAAGGCGACCCCGTTCAAGGAGTTAGTGTAAAGGCGGCCAGCTTGGCCACCGGTGTTCCGG
 580 590 600 610 620 630 640
 650 660 670 680 690 700 710
 TAACCACTCACACGGCAGCAAGTCAGCGCGATGGTGAGCGAGGCAGGCCATTTGAGTCCGAAGCTTG
 TAACCACTCACACGGCAGCAAGTCAGCGCGATGGTGAGCGAGGCAGGCCATTTGAGTCCGAAGCTTG
 650 660 670 680 690 700 710
 720 730 740 750 760 770 780
 AGCCCTCACGGGTTTGTATTGGTCACACCGATGATACTGACGATTGAGCTATCTCACCGCCCTGCTGCGCG
 AGCCCTCACGGGTTTGTATTGGTCACACCGATGATACTGACGATTGAGCTATCTCACCGCCCTGCTGCGCG
 720 730 740 750 760 770 780
 790 800 810 820 830 840 850 860
 GATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAGATAATGCGAGTGCATCACCGC
 GATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAGATAATGCGAGTGCATCACCGC
 790 800 810 820 830 840 850 860
 870 880 890 900 910 920 930
 TCCTGGGCATCCGTTGCGCAAAACACGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGCTACATGAAAC
 TCCTGGGCATCCGTTGCGCAAAACACGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGCTACATGAAAC
 870 880 890 900 910 920 930
 940 950 960 970 980 990 1000
 AAATCCTCGTTCGAATGACTGGCTGTTGGGTTTCGAGCTATGTCACCAACATCATGGACGTGATGGATC
 AAATCCTCGTTCGAATGACTGGCTGTTGGGTTTCGAGCTATGTCACCAACATCATGGACGTGATGGATC
 940 950 960 970 980 990 1000
 1010 1020 1030 1040 1050 1060 1070
 GCGTGAACCCCGACGGGATGCCCTTCATTCCACTGAGAGTGATCCATTCTACGAGAGAAAGGGCGTCCCACA
 GCGTGAACCCCGACGGGATGCCCTTCATT-CACTGAGAGTGATCCATTCTACGAGAGAAAGGGCGTCCCACA
 1010 1020 1030 1040 1050 1060 1070
 1080 1090 1100 1110 1120 1130 1140
 GGAAACGCTGCCAGGCATCACTGTGACTAACCGGCGCGTTCTGTGTCACCGACTTGCCGTGCATGACGCC
 GGAAACGCTGCCAGGCATCACTGTGACTAACCGGCGCGTTCTGTGTCACCGACTTGCCGTGCATGACGCC
 1080 1090 1100 1110 1120 1130 1140
 1150 1160 1170 1180 1190 1200 1210 1220
 ATCTGGATCCTTCACCGCAGCGGCCACTATTCCCCGTCAAGATAACCGAACGATGAAAGTCGCGCATCGATCGA
 ATCTGGATCCTTCACCGCAGCGGCCACTATTCCCCGTCAAGATAACCGAACGATGAAAGTCGCGC---ATCGA
 1150 1160 1170 1180 1190 1200 1210
 1230 1240 1250 1260 1270 1280 1290
 TAGGCATCTCAATGTRGATCAGGGCTGCGCGTCAAGGAAAGCGCGGTGCCACCCCTGTCGATAGTCTTGAGGGGA
 TAGGCATCTCAATGTRGATCAGGGCTGCGCGTCAAGGAAAGCGCGGTGCCACCCCTGTCGATAGTCTTGAGGGGA
 1220 1230 1240 1250 1260 1270 1280
 1300 1310 1320 1330
 CGGTAGCGACGACCGTCTTCTGCTGAG

CGGTAGCGACGACCGCTGCTTTCTGCTAGCTGAG
1290 1300 1310 1320 X

3. LOW344-FIG1. SEQ

FVBOPD Flavobacterium sp. parathion hydrolase gene, compl.

LOCUS FVBOPD 1883 bp ds-DNA BCT 15-JUN-1990
DEFINITION Flavobacterium sp. parathion hydrolase gene, complete cds.
ACCESSION M29593
KEYWORDS parathion hydrolase.
SOURCE Flavobacterium sp. (strain ATCC 27551) DNA, clone pPDL2.
ORGANISM Flavobacterium sp.
Prokaryota; Bacterial; Gracilicutes; Scotobacteria; Neisseriaceae;
Flavobacterium; sp.
REFERENCE 1 (bases 1 to 1883)
AUTHORS Mulbry, W. W. and Karns, J. S.
TITLE Parathion hydrolase specified by the Flavobacterium opd gene:
Relationship between the gene and protein
JOURNAL J. Bacteriol. 171, 6740-6746 (1989)
STANDARD simple staff_entry
FEATURES Location/Qualifiers
misc_signal 312..317
/note="-35 region"
misc_signal 334..339
/note="-10 region"
misc_binding 408..411
/note="ribosome binding site"
CDS 419..1516
/note="parathion hydrolase"
BASE COUNT 372 a 457 c 477 g 347 t
ORIGIN 1 bp upstream of BamHI site.

Initial Score = 589 Optimized Score = 1281 Significance = 261.10
Residue Identity = 96% Mismatches = 1304 Mismatches = 14
Gaps = 28 Conservative Substitutions = 0

	X	10	20	30	40	50	60
	CTGCAGCCTEACTCGGCACCAAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGC						
	CGGTTCAAGATCTGCAGCCTEACTCGGCACCAAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGC						
350	360	370	380	390	400	410	
	70	80	90	100	110	120	130
	ATGCAAACGAGAAGGGTTGTGCTCAAGTCTGCG---GCCGCAGGAACTCTGCTGGGGGGCCTGGCTGGGTGC						
	ATGCAAACGAGAAGGGTTGTGCTCAAGTCTGCGCCGCCAGGAACTCTGCTGGGGCCTGGCTGGGTGC						
420	430	440	450	460	470	480	490
	140	150	160	170	180	190	200
	GCGA-CGTGGCTGGATCGATCGCACAGCGATCGATCAATA-CGTGCGC-GTCCTATCACAAATCTCTGAA						
	GCGAGCGTGGCTGGATCGATCGCACAGCGATCGATCAATAACCGTGCGCGGGTCCTATCACAAATCTCTGAA						
500	510	520	530	540	550	560	
	210	220	230	240	250	260	270
	GCGGGTTTCACACTGACTCACGAGGACATCTGCGGCAGCTCGGCAGGGATTCTTGCCTGGGCCAGAGTTC						
	GCGGGTTTCACACTGACTCACGAGGACATCTGCGGCAGCTCGGCAGGGATTCTTGCCTGGGCCAGAGTTC						
570	580	590	600	610	620	630	
	280	290	300	310	320	330	340
	TTCGGTAGCCGCAAAGCTCTAGCGA-AAGAAGGCTGTCAGAGGGATTG---CGCGGCCAGAGCAGGGCTGGCGTGCAG						
	TTCGGTAGCCGCAAAGCTCTAGCGA-AAGAAGGCTGTCAGAGGGATTGCGCGCCAGAGCAGGGCTGGCGTGCAG						

640 650 660 670 680 690 700
350 360 370 380 390 400 410
ACGATTGTCGATGTGTCGACTTTCGATATCGCTCGCGACGCTCAGTTATTGGCCGAGGTTTCGCGGGCTGCC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACGATTGTCGATGTGTCGACTTTCGATATCGCTCGCGACGCTCAGTTATTGGCCGAGGTTTCGCGGGCTGCC
710 720 730 740 750 760 770
420 430 440 450 460 470 480
GACGTTCATATCGTGGCGCGACCGGCCTTCGACCCGCCACTTCGATGCGATTGAGGTATGTAGAG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GACGTTCATATCGTGGCGCGACCGGCCTTCGACCCGCCACTTCGATGCGATTGAGGAGTGTAGAG
780 790 800 810 820 830 840 850
490 500 510 520 530 540 550
GAACTCACACAGTTCTCCTGCGTGAATTCATATGGCATCGAAGAGACACCGGAATTAGGGCGGGCATTATC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GAACTCACACAGTTCTCCTGCGTGAATTCATATGGCATCGAAGAGACACCGGAATTAGGGCGGGCATTATC
860 870 880 890 900 910 920
560 570 580 590 600 610 620
AAGGTCGCGACCAACAGGCAAGGCGACCCCGCTTCAGGAGTTAGTGTAAAGGCGGCCGGCCAGCTG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AAGGTCGCGACCAACAGGCAAGGCGACCCCGCTTCAGGAGTTAGTGTAAAGGCGGCCGGCCAGCTG
930 940 950 960 970 980 990
630 640 650 660 670 680 690 700
GCCACCGGTGTTCCGGTAACCACTCACACGGCAGCAAGTCAGCGCGATGGTGAGCGAGGCAGGCCATT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GCCACCGGTGTTCCGGTAACCACTCACACGGCAGCAAGTCAGCGCGATGGTGAGC-AG-CAGGCCGCCATT
1000 1010 1020 1030 1040 1050 1060
710 720 730 740 750 760 770
TTGAGTCCGAA-GCTTGAG-CCTCAGCGGTTGTATTGTCACAGCGATGATACTGACGATTGAGCTATC
||||||||||||||||||||||||||||||||||||||||||||||||||||
TTGAGTCCGAAAGGCTTGAGCCCTCACCGGTTGTATTGTCACAGCGATGATACTGACGATTGAGCTATC
1070 1080 1090 1100 1110 1120 1130
780 790 800 810 820 830 840
TCACCGCCCT-GCT--GCGCGGATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAG
||||||||||||||||||||||||||||||||||||||||||||
TCACCGCCCTCGCTGCGCGCGGATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAG
1140 1150 1160 1170 1180 1190 1200
850 860 870 880 890 900 910
ATAATGCGAGTGCATCACCGCTCCTGGCATCCGTTTCGTTGGCAAACACGGGCTCTCTGATCAAGGCGCTCA
||||||||||||| | ||||||||||||||||||||||||||||||
ATAATGCGAGTGCATCACCGCTCCTGGCATCCGTTTCGTTGGCAAACACGGGCTCTCTGATCAAGGCGCTCA
1210 1220 1230 1240 1250 1260 1270 1280
920 930 940 950 960 970 980
TCGACCAAGGCTACATGAAACAAATCCGTTTCGAAATGACTGGCTGTTGGGTTTCGAGCTATGTCACCA
||||||||||||||||||||||||||||||||||||||||
TCGACCAAGGCTACATGAAACAAATCCGTTTCGAAATGACTGGCTGTTGGGTTTCGAGCTATGTCACCA
1290 1300 1310 1320 1330 1340 1350
990 1000 1010 1020 1030 1040 1050
ACATCATGGACGTGATGGATCGCGTGAACCCCGACCGGGATGGCCTTCATTCCACTGAGAGTGTGATCCCATT-C
||||||||||||||||||||||||||||||||||||||||
ACATCATGGACGTGATGGATCGCGTGAACCCCGACCGGGATGGCCTTCATTCCACTGAGAGTGTGATCCCATTCC
1360 1370 1380 1390 1400 1410 1420
1060 1070 1080 1090 1100 1110 1120
TACGAGAGAAGGGCGTCCCACAGGAAACCGCTCCAGGCATCACTGTGACTAACCCGGCGCGGGTTCTGTGTC
||||||||||||||||||||||||||||||||||||||||
TACGAGAGAAGGGCGTCCCACAGGAAACCGCTCCAGGCATCACTGTGACTAACCCGGCGCGGGTTCTGTGTC
1130 1140 1150 1160 1170 1180 1190

1430	1440	1450	1460	1470	1480	1490
1130	1140	1150	1160	1170	1180	1190
CCGA-CTTGC---CGTGCATGACGCCATCTGGATCCTTCACGCAGCGGCCACTATTCCCCGTCAAGATAACC						
1500	1510	1520	1530	1540	1550	1560
1200	1210	1220	1230	1240	1250	1260
GAACGATGAAGTCGCGCATCGATCGATAGGCATCTCAATGTGATCAGGGCTGCCACCTCCAAAGGCCGGTGG						
1570	1580	1590	1600	1610	1620	1630
1270	1280	1290	1300	1310	1320	X
CCACCCCTGTCGATAGTCTTGAGGGACGGTAGCGACGACCGTGTCTTCGTGAACTGCAG						
1640	1650	1660	1670	1680	1690	X
CCACCCCTGTCGATAGTCTTGA-GGAC-GTAGGGCACACCGTGCTTTG--GAACTGCAG						

4. LOW344-FIG1. SEQ

TRN21TNPA Transposon Tn21 tnpA gene for transposase.

LOCUS	TRN21TNPA	3176 bp ds-DNA	BCT	15-MAR-1988				
DEFINITION	Transposon Tn21 tnpA gene for transposase.							
ACCESSION	X04891							
KEYWORDS	tnpA gene; tnpR gene; transposase; transposon.							
SOURCE	Transposon Tn21.							
ORGANISM	Transposon Tn21							
REFERENCE	Prokaryota; Bacteria; Transposon Tn21.							
AUTHORS	1 (bases 1 to 3176)							
TITLE	Ward, E. and Grinsted, J.							
JOURNAL	The nucleotide sequence of the tnpA gene of Tn21							
STANDARD	Nucleic Acids Res. 15: 1799-1806 (1987)							
COMMENT	simple automatic							
FEATURES	[1] enum. 1 to 3176.							
CDS	Location/Qualifiers							
	(1..54							
	/note="tnpR gene"							
CDS	57..3023							
	/note="transposase (AA 1-988)"							
BASE COUNT	652	a	1057	c	915	g	552	t
ORIGIN								

Initial Score = 87 Optimized Score = 606 Significance = 15. 98
 Residue Identity = 52% Matches = 761 Mismatches = 472
 Gaps = 212 Conservative Substitutions = 0

X	10	20	30	40	50	
CTGCAGCCTGACTCGGCA--CCAG-TCGC--TGCAAG----CAGAGTCGTAAGCAATCGCAA						
70	80	90	100	110	120	130
TCCATCCTGTCCGCCGCC-GAGCGGGAAAGCCTGCTGGCGTTGCCGGACTCCAAGGGACG-ACCTGATC-CGA						
60	70	80	90	100	110	120
GGGGGCAGCATGCAAACGAGA-AGGGTTGT-GCTCA--AGTCTGCGGC-CG---CAGGAACCTGCTGGCG						
140	150	160	170	180	190	200
CATTACA-CATTC-AACGATAACCGAACCTCTCGATCATCCGACAGCGGGCGCGGGCCAGCCAATCGGCTGGGCT						
120	130	140	150	160	170	180
GCCTGGCTG--GGTGCAGCGACGTC-GCTGGATCGATCGGCA--CAGGCGATCGGATCAA-TA-CGTGCGCGT						
210	220	230	240	250	260	270
TOCGGG-TGCAGCTCTGTTACCTGCGCTTCCCG-GCGTCATCCTGGCGTC-GATGAACTACCGTTCCGC						

190 200 210 220 230 240 250
 CCTATCACAACTCTCTGAAGCGGGT /TCACACTGACTCAGGAGACATCTGCGGCAGCTCGGCAGGATTCTG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CCT-----TGT-TGAAGCCGGTUGCGGAC-CAGUCA-AAGGTGGCGTCGAAAGCT-GGAACGAGTACGG
 280 290 300 310 320 330

 260 270 280 290 300 310 320
 CGTGCTTGGCCAGAGTTCTCGGTAGCGCGCAAAGCTCTAGCGGAAAAGGCTGTGAGAGGGATTGCG-----C
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CCAGCGGGAGGCAGA-CCCGCGCGAGCACCTGAG--CGAGCTGCAA--CCGTGTTGGTTCCGGCCCTC
 340 350 360 370 380 390 400

 330 340 350 360 370
 GCCA-GAG-----CGGCTGG-CGTGCGAACGATTGTCATGTGCGA--CTTTC-GATATCGG--TCGCG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ACCATGAGCCATTACCGGGCAGGCCGATGCTGACCGAGCTGGCGATGCAAACCGACAAAGGCATCGTG
 410 420 430 440 450 460 470

 380 390 400 410 420 430 440
 -ACGTCAGTTATTGGCCGAGGTTTCGCGGGCTGCCGACGTTCATATCGTGGCGGCGACCGGGCTTGTGGTTC
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CTGGCCAGCGCTTGATCG-GGCACCTCGGGCGCAGTCGGTCAT-TC-TGCC--CGCCCTAACGCGC-TC
 480 490 500 510 520 530 540

 450 460 470 480 490 500 510
 GA-CCCGCCACTTCGATGCGATTGAGGTATAGAGGAACTCACACAGTTCTTC--CTGCGTG--AGATTG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GAGCGGGCGAGTGCGGAGGCAGA-TCACCCGTG---CTAACCGGGCGA--TCTACGACGCCTGGCGAAC
 550 560 570 580 590 600

 520 530 540 550 560 570 580
 AAT-ATGGCATCGAACACCGGAATTAAAGGCGGCCGGGCGAGCTTGC--CA---CCGGTGTTCCGGTAACCAC
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 ACTGGCGGACGCGCA-TCGCCGCCCTCGACGATC-TGCTCAAGCGCCGGGACAAC-GGCAAGACGACCTG
 610 620 630 640 650 660 670

 590 600 610 620 630 640 650
 CTTTCAGGAGTTAGTGTAAAGGCGGCCGGCGAGCTTGC--CA---CCGGTGTTCCGGTAACCAC
 ||| ||| ||| ||| ||| ||| ||| |||
 GTT---GGCTTGGTTGCGCCAGTC-TCCGGCCAAGCAAATTGCGGGCATATGCTGGAACACATCGAACGCC
 680 690 700 710 720 730 740

 660 670 680 690 700 710
 TCA---CACGGCAGCAAGTC-AGC-GC-GATGGTGAGCGA--GGCAGG-CCGCCATTGGTGAAGT-CCGAAGC
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 TCAAGGCATGGCAGGCACTCGATCTGCCAACCGGCATCGAGCGGCTGGTTACCAAGAACCGCCTGCTCAAGA
 750 760 770 780 790 800 810

 720 730 740 750 760 770 780
 TTGAGCCCTCACGGGTTGTATTGGTACAGCGATGATACTGAC--GATTTGAGCTATCTCACCGCCCTGCT
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 TT--GCCCGCGAGGGCGGCCAGATGACAC-CCGCCGA-CCTGGCCAATTGAGC---CGCAACGGC--GCT
 820 830 840 850 860 870

 790 800 810 820 830 840 850
 GCGCGGATAACCTCATCGGTCTAGAACACATCCCGCACAGTGCGATTGGTCTAGAACATAAT--GCGA-GTGC
 ||| ||| ||| ||| ||| ||| ||| ||| |||
 ACGCCACT--CTCGTGGCGCT-GGCCAC--CGAGGGCATGGCCACCGTCACCGACGA-AATCATCGACCTGC
 880 890 900 910 920 930 940

 860 870 880 890 900 910
 ATCACCGC-TCCCTGGC-CATCGGT---CG-TGCGAACACCGGC-TC---TCTTGATCAAGGCGCTCATCG
 ||| ||| ||| ||| ||| ||| ||| |||
 ACGACCGCATCCTGGGTAAGCTGTTAACCGCTGCCAAGAATAAGCATCAGCAGCAGTTCCAGGGCG-TCA--G
 950 960 970 980 990 1000 1010

920	930	940	950	960	970	980
ACCAAGGCTACATGAAACAAATCCCTCGT TTGAAATG-ACTGGCTGTTGGGTTTCGAGCT-AT-GTCACCA						
1020	1030	1040	1050	1060	1070	
990	1000	1010	1020	1030	1040	
A-CA-TCATGGACGTGATG-GATCG-CG TGAACCCCGACGGGATGGCCT-TCATTCC---ACTGAGAGGTGAT						
1080	1090	1100	1110	1120	1130	1140
1050	1060	1070	1080	1090	1100	1110
CCCATTCTACGAG-AGAAGGGCGTCCCACAGGAAACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGGTT						
1150	1160	1170	1180	1190	1200	1210
1120	1130	1140	1150	1160	1170	1180
CTGTGTACCGACT-TGCCG---TGCATGACGCCAT--CTGGATCCTTC-CACGCAGCGGCCACTATTCCCC						
1220	1230	1240	1250	1260	1270	
1190	1200	1210	1220	1230	1240	1250
GTCAAGATAACCGAACGATGAAGTCGCGCATC-GATCG---ATAGGCATCTTCAATGTGATCAGGGCTGCCACC						
1280	1290	1300	1310	1320	1330	1340
1260	1270	1280	1290	1300	1310	
---TCCAAAG-CCGG---TGCCTCA---CCCTCTCG-ATAGTCTTGGGGACGGTAGCGACGACCGTGCTT---T						
1350	1360	1370	1380	1390	1400	1410
1320	X					
TCGTG---AACTGCAG						
1420	1430					
CCGGGCCTACTACGAACGTGCGCG						

5. LOW344-FIG1. SEQ

X17379 *Sorghum vulgare* mRNA for phosphoenolpyruvate invol

LOCUS X17379 3147 bp UNA 15-SEP-1990
 DEFINITION Sorghum vulgare mRNA for phosphoenolpyruvate involved in C4
 photosynthesis (EC 4.1.1.31).
 ACCESSION X17379
 KEYWORDS
 SOURCE
 ORGANISM Unknown
 Unclassified.
 REFERENCE 1 (bases 1 to 3147)
 AUTHORS Cretin,C. , Keryer,E. , Tagu,D. , Lepiniec,L. , Vidal,J. and Gadal,P.
 TITLE Complete cDNA sequence of sorghum phosphoenolpyruvate carboxylase
 involved in C4 photosynthesis
 JOURNAL Nucleic Acids Res. 18: 658-658 (1990)
 STANDARD unannotated staff_entry
 REFERENCE 2 (bases 1 to 3147)
 AUTHORS Cretin,C. D.
 JOURNAL Unpublished (1989) see COMMENT for author address
 STANDARD unannotated staff_entry
 BASE COUNT 675 a 946 c 915 g 611 t

ORIGIN

Initial Score = 80 Optimized Score = , 597 Significance = 12.71
Residue Identity = 51% Matches = 740 Mismatches = 490
Gaps = 203 Conservative Substitutions = 0

130	140	150	160	170	180	190	
GCTGGGTGCGCGACGTGGCTGGATCGATCG-GCACAGGCGATCGGATCAATACTGCGCGTCCTATCACAAAT							
GC-GCCCTCG-GACGT-GCTCG-CCG-TCGAACTCCTCGC-AGCG---CGAGATGTG-GCATTGCCAGCGGT	1600	1610	1620	1630	1640	1650	1660

200	210	220	230	240	250
CTC--TGAAGCGGGTTT	CACACTGACTCACGAGGACATCTGC	GGCAGCTCGGCAGGATTCTTG	-----	CGTG	
CCCCGTGGTGCCGCTGTTGA	--GAGGCTGGCCGAC	--CTGCAGGCAGGCGCGCG	--TCCGTGGAGAAGCT		
1670	1680	1690	1700	1710	1720

260	270	280	290	300	310	320	
CTTGGCCAGAGTTCTCGGTAGCCGCAA--AGCTCTAGCGGAAAAGGCTGTGAGAGGGATTGCGCGCCAGAGC							
CTTCTCCACTG-ACT--GGTA-CATCAACCACATC-AACGGCAA--GCAG-CAGGTGATGGTCGGCTACTCC	1730	1740	1750	1760	1770	1780	1790

330	340	350	360	370	380	390			
GGCTGGCGTGC	GAACGATTGTCG	-ATGT	-GTCGACTT	CGATATCGGT	-CGCGACGT	CAGTTA-TTGGCCG			
GA	CTCGGG	CAAGG	ACGCCG	GGCCG	CCTG	TGACG	TGGCGC	CAGGAGGAGA	TGGCCA
1800	1810	1820	1830	1840	1850	1860			

```

530      540      550      560      570      580      590
AAGACACCGGAATTAGGGCGGGCATTATCAAGGTCCGACCAAGGCAAGGGGACCC---CCTTCAGGAGT
|  ||  |  ||  ||  ||  |  |  ||  ||  ||  |  |  ||  ||  ||  |  ||  ||  |
TGGTGAC--GGTGAGGGCGAG-GTCATCGA&TTCATGTTC---GGGGAGGGAGAACCTGTGCTTCCAGTCTC
  530  5410  5500  5600  5700  5810  5950

```

2

-AG

1 1
1 1

AGCGGAGT

6. LOW344-FIG1. SEQ

NEUTRP1 n. crasse trifunctional tryptophan biosynthesis gen

LOCUS NEUTRP1 2750 bp DNA PLN 01-AUG-1983
 DEFINITION n. crassa trifunctional tryptophan biosynthesis gene trp-1.
 ACCESSION JO1252
 KEYWORDS multifunctional enzyme.
 SOURCE neurospora crassa.
 ORGANISM Neurospora crassa
 Eukaryota; Plantae; Thallobionta; Eumycota; Ascomycotina;
 Pyrenomycetes; Sordariales; Sordariaceae; Neurospora; crassa.
 REFERENCE 1 (bases 1 to 2750)
 AUTHORS Schechtman, M. G. and Yanofsky, C.
 TITLE structure of the trifunctional trp-1 gene from neurospora crassa
 and its aberrant expression in escherichia coli
 JOURNAL J. Mol. Appl. Genet. 2, 83-99 (1983)
 STANDARD simple staff_review
 BASE COUNT 572 a 821 c 727 g 630 t
 ORIGIN

Initial Score = 85 Optimized Score = 594 Significance = 11.62
Residue Identity = 51% Matches = 738 Mismatches = 487
Gaps = 213 Conservative Substitutions = 0

1190	1200	1210	1220	1230	1240	1250
TCAAGATACCGAACGATGAAGTCGCGCATCGATCG-ATAGGCA--TCTTCATGTGATCAGGGCTGCCACCT						
1460	1470	1480	1490	1500	1510	1520
TCAAGATGCTCGAGTATGA---GCTCCCGAGCGCTATAACAGTACTCCTTGT-CTCTCGGCATGGAGC-						
1260	1270	1280	1290	1300	1310	
CCAAAGCCGGTGGCC--ACCCCTGTCGATA-GTCTTGGGGGA-CGGTAGCGACGACCGTGCTTT---CGTG						
1530	1540	1550	1560	1570	1580	1590
CCCTAGTCGAGGTCCAGAACACCGAGGAGATEGCCACAGCCATCAAGCTCGCG-CCAAGGTTATCGGGCTC						
1320	X					
AACTGCAG						
1600	X	1610				

7. LOW344-FIG1. SEQ

BLYAMYAA Barley alpha-amylase type A isozyme mRNA, complete

LOCUS BLYAMYAA 1588 bp ss-mRNA PLN 11-NOV-1985
 DEFINITION Barley alpha-amylase type A isozyme mRNA, complete cds.
 ACCESSION J01236
 KEYWORDS alpha-amylase; amylase.
 SOURCE Barley (Hordeum vulgare L. cv. Himalaya 1979 crop) aleurone cell
 stimulated with gibberellic acid, cDNA to mRNA, clone E.
 ORGANISM Hordeum vulgare
 Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
 Commelinidae; Cyperales; Poaceae; Hordeum; vulgare.
 REFERENCE 1 (bases 1 to 1588)
 AUTHORS Rogers, J. C. and Milliman, C.
 TITLE Isolation and sequence analysis of a barley alpha-amylase cDNA
 clone
 JOURNAL J. Biol. Chem. 258, 8169-8174 (1983)
 STANDARD full staff_review
 COMMENT [1] suggests that there are two alpha-amylase genes in barley
 aleurone cells (types A and B), and that expression of these genes
 is affected differently by gibberellic acid. It is likely that
 alpha-amylase contains a signal peptide and a mature peptide. The
 latter would start at about position 172 (by comparison with
 alpha-amylase type B). A poly-A signal is located at positions
 1557-1562.
 FEATURES Location/Qualifiers
 mRNA 1.. 1588
 /note="a-amyl mRNA"
 CDS 97.. 1413
 /note="alpha-amylase type A, EC 3.2.1.1"
 BASE COUNT 344 a 484 c 480 g 280 t
 ORIGIN 95 bp upstream of NcoI site; chromosome 1.

Initial Score = 107 Optimized Score = 592 Significance = 10.89
 Residue Identity = 50% Matches = 718 Mismatches = 533
 Gaps = 171 Conservative Substitutions = 0

X	10	20	30	40	50	
CTGCAGCCTGACTCGGC-ACCACTCGCTGCAAGCAGAGTCGT--AAGCA-ATCGCAAGGGGG						
160	X	170	180	190	200	210
GGTTGGCGTCCGGCCACC-AAGTCTCTTTCAGGGCTCAACTGGGAGTCGTGGAAGCAGAGCGGGCGGGTGG						
60	70	80	90	100	110	120
--CAGCATGCAAACGAGAAGGGTTGTCGTCAGATCTCGCGGCCAGGAACTC-TGCTGGGCGGCCTGGCTGG						
110	120	130	140	150	160	170

TACAAACATGATGATGGCAAGTCGACCGACATCGGCTGCCACCGCCG
230 240 250 260 270 280 290
130 140 150 160 170 180 190
GTGCGCGACGCTGGCTGGATCGATCGCACAGGC--GATCGGATCAATACTGCGCGTCCTATCACAATCTCT
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TCGCACTCCGT--CTCCAACGAAGGGTTACATGCGCTGGTCGG--CTGTACGACATCGACGCGTC-CAAGTACG
300 310 320 330 340 350 360
200 210 220 230 240 250 260
GAAGCGGGTTTCACACTGA-CTCACG-AGGACATCTGCG-GCAGCTCGGCAGGATTOTTGC-GTGCCTGGCC
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GCAACGCG-----GCGGAGCTCAAGTCGCTCATCGCGCGCTCACGGCAAG-GGCAGTCAGGCCATCGCC
370 380 390 400 410 420
270 280 290 300 310 320
AGAGTTCTTCGGTAGCCGCAAAGCTCTAGCG--GAAAAGGCTGTGAGAGG-AT-TGC-GC---GCCAGAGC
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
-GACATCGTCATCAAC--CACCGCTGCCCGACTACAAGGATAGCCGCGGCATCTACTGCATCTCGAGGGC
430 440 450 460 470 480 490
330 340 350 360 370 380 390
GGCTGGCGTGCGAACGATTG--TCGA-TGTGTCGACTTTGATATCGGTCGCGACGTCAGTTATTGGCCGA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GGC--ACCTCCG-ACGGCCGCGCTCGACTGGGGC-CCCCACATGATCTGTCGCGACGACACCAAATACTCCGA
500 510 520 530 540 550 560
400 410 420 430 440 450 460
GGTTTCGCGGGCTGCCGACGTTCATATCGTGGCGGCGAC--CGGCTTGTGGCTCGAC-CCG-CCACTTTCGA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
----TGGC--ACCGCAAACCTCGACACC--GGAGCCGACTTCGCCGCCGCGACATCGACCCAC-CTCAA
570 580 590 600 610 620
470 480 490 500 510 520
TGCATTGAGGTATGTAGAG-GAACTCACACAGT---TCTTCCTGCGTGAGATTCAATATGGCATCGA---
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
---CGACCG-GGT--CCAGCGCGACGCTCAAGGAGTGGCTCCTCTGGCTCAAGAGCGACCTCGGCTCGACGCG
630 640 650 660 670 680 690
530 540 550 560 570 580 590
AGACACC-GGAATTAGGGCGGGCATTATC-AAGGTCGCGACCGACAG-GCAAGGCGAC-CCCCTTCAGGAG-
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
TGGCGCCTTGACTTCGCTAGGGCTACTCGCGGGAGATGCCAAGGTGTACATCGACGGCACATCCCCGAGC
700 710 720 730 740 750 760
600 610 620 630 640 650 660
TTAGTGTAAAGGCGGCCGCCGGGCCAGGCTTGGCCACCGG-TGTTCCGGTAACCACTCACACG-GCAGCAA
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CTCGCCGT---GGCCGAGGTGTGGGACAAATATGGCCACCGGGCGGACGGCAAGGCCAACTACGACCAAGGAC
770 780 790 800 810 820 830
670 680 690 700 710 720 730
G-TCAGCG--CGATGGTGAGCGAGGCAGGCCATTGTTGAGTCGAAGGCTTGAGCCCTCACGGGTTGTA
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
GCGCACCGGCAGAATCTG-GTGAACCTGCGTGGACAAGGTGGCGGCCCTCGGCAGGCATGGTGTGCGA
840 850 860 870 880 890 900
740 750 760 770 780 790 800
TTGGTCAC-AGCGATG--ATACTG-ACGATTGAGCTATCTCACCGCCCTGCTGCGCGGATAACCTCATCGGT
||| ||| ||| ||| ||| ||| ||| ||| |||
CT--TCACGACCAAAAGGGATACTCAACGCT----GCCGT-GGAGGG-CGAGCT--GTGGA-GGCTGATCGAC
910 920 930 940 950 960
810 820 830 840 850 860 870
CTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAGATAATGCGAGTGCGATCACCGCTCCTGGGCATC--C
||| ||| ||| ||| ||| ||| ||| |||

CCGCAGGGGAAGGCCCCGGGATGTA-TGGATGG-TGGCGGCCAAGGCCGCC-TCGTCGACAACCAC
 970 980 990 1000 1010 1020 1030
 880 890 900 910 920 930
 GTTCGTGGC-AAACACGGGCTCTCTTGTATCAAGGGCTCATCGACCAAGGC-TACA-TGAAAC----AAATC
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GATACAGGCTCCACGCAGGC-CATGTGGCCATTCCCTC--CGACAAGGTATGCAGGGCTACGCGTACATC
 1040 1050 1060 1070 1080 1090 1100
 940 950 960 970 980 990 1000
 CTCGTTTCGAATGACTGGCTGTTGGGTT-TTCGAGCTATGTCACCAACATCATGGACG-TGATGGATCGCG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CTCACCCACCCCGGCATCCCATGCATCTTACGA-CCATTCTTCAAC---TGGGGGTTAAGGA---CC
 1110 1120 1130 1140 1150 1160 1170
 1010 1020 1030 1040 1050 1060 1070
 TGAACCCCGACGGGATGGCCTTCATTCACAGAGTGATCCCATTCTACGAG----AGAAGGGCGTCCC
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 AGATCGCGCGCTGGTGGCGATCAGGAAGCGCAACGGCATTACAC-GGCGACGAGCGCTCTGAAGATCCTCATG
 1180 1190 1200 1210 1220 1230 1240
 1080 1090 1100 1110 1120 1130 1140
 CAGGAA-ACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGGGTTCTGTGTCACCGACTTGCCGTGCATG-A
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CACGAAGGAGATGCC-TACGTGCCGAGA-TA-GACGGCAAGGTGGTG-GTGA-AGA-TCG-GGTCCAGGTA
 1250 1260 1270 1280 1290 1300
 1150 1160 1170 1180 1190 1200 1210
 CGCCATCTGGATCCT--TCCACGCAGCGGCCACTATTCCCCG--TCAAGATACCGAACGATGAAGTC-GCGC
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 CGACGTCGGGGCGGTGATCCCGGC--CGGGTTCGTGCACCTCGGCACACGGCAACG-ACTACGCCGTCTGGGA
 1310 1320 1330 1340 1350 1360 1370
 1220 1230 1240 1250 1260 1270
 ATCGATCGAT----AGGCA--TCTTCAA-TGTGATCAGGGCTGC-CACCTCC-AAAGCCGGT-GGCCACCCC
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 GAAGAACGGTGCACGGCAACACTACAACGGAGCTGAAGTCTGCACTGATCCGTATTGAGCATGAA
 1380 1390 1400 1410 1420 1430 1440
 1280 1290 1300 1310 1320 X
 TGTC--GATAGT-CTTGAGGGAC----GGTACCGACGACCG---TGCTTTTC-GTGAACGTGCAG
 ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 TTTCTGA-AGTACATGATTCACTTCTGGTATTCAAG-CGGATATGATTAACATGTATACCTGTACCCAAA
 1450 1460 1470 1480 1490 1500 1510
 AT
 1520

8. LOW344-FIG1. SEQ

PDUMER Plasmid pDU1358 (from *S. marcescens*) mercurial resi

LOCUS	PDUMER	2153 bp ds-DNA	BCT	15-MAR-1988
DEFINITION	Plasmid pDU1358 (from <i>S. marcescens</i>) mercurial resistance (mer) operon encoding organomercurial lyase (merB), mercury resistance protein (merD), complete cds, and mercury reductase (merA), 3' end. M15049			
ACCESSION				
KEYWORDS	antibiotic resistance; mercuric reductase; mercury resistance; organomercurial lyase; transport protein.			
SOURCE	Plasmid pDU1358 (multi-antibiotic resistance IncC incompatibility) DNA (from <i>Serratia marcescens</i>), clone pHG6.			
ORGANISM	Plasmid pDU1358 Prokaryota; Bacteria; Plasmid pDU1358.			
REFERENCE	1 (bases 1 to 2153)			
AUTHORS	Griffini, H. G., Foster, T. J., Silver, S. and Misra, T. K.			
TITLE	Cloning and DNA sequence of the mercuric and organomercurial			

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 84, 3112-3116 (1987)
STANDARD full staff_review
COMMENT Computer-readable sequence for [1] kindly provided by S. Silver,
25-MAR-1987.

FEATURES Location/Qualifiers
CDS <1..359
/note="mercuric reductase (merA; AA at 3)"
CDS 374..1012
/note="organomercurial lyase (merB)"
CDS 1124..1489
/note="mercury resistance protein (merD)"
BASE COUNT 391 a 641 c 680 g 441 t
ORIGIN Unreported.

Initial Score = 95 Optimized Score = 580 Significance = 10.17
Residue Identity = 50% Matches = 727 Mismatches = 511
Gaps = 184 Conservative Substitutions = 0

X 10 20 30 40 50
CTGCAGCCTGACTC-GGCACCAGTCGCTGCAAG-CAG--AGTC-GTAAGCAATCGCA--AGG
|| || || || || || || || || || || || || || || || || || || || || || || || ||
TGGCGGTGTCCTGGTATTGCCGAGGAAGCAGCCGACGTTCGTCAGTCCTCTGTTGCCATGTACATTCT
800 X 810 820 830 840 850 860

60 70 80 90 100 110 120
GGGCAGC-ATGCAAACG--AGAAGGGTTGTGCTCAAGTCTGCGGCCGAGGAACCTCTGCTGGGCGGCCTGGC
|| || || || || || || || || || || || || || || || || || || || || || || || || || ||
TTGCATCTGTCCCGACGGCGGAAGACTGGGCCTCCAAGCATCAAGGATTGGAA---GGATTGGCGATC-GTC
870 880 890 900 910 920 930

130 140 150 160 170 180 190
TGGGTGCGCGACGTGGCTGGATC-GATCGGCACAGGCGATCGGATCAATACGTG-CGCGTCCTATCACAAATC
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
AGTGT---CCACGAGGCT--TTCGGCTTGGGCCAGGAG-TTTAATCGACATCTGTTGCAGACCATGTC-ATC
940 950 960 970 980 990

200 210 220 230 240 250 260
TCTGA-AGCG-GGTTTCACACTGACTCACGAGGACATCTGCGGCAGCTCGGCAGGATTTC-TTGCCTGC---T
|| || || || || || || || || || || || || || || || || || || || || || || || || || ||
TAGGACACCGTGATCGGATATCGACCCA--ATG--TTCTACGGCACCGGCATCGGATTCGCAGCGCGCGGAT
1000 1010 1020 1030 1040 1050 1060

270 280 290 300 310 320
TGGCCAGAGTTCTCGGTAGCCGCAAAGC--TCTAGCG--GAAAAG--GCTGTGAGAGGGATTGCGCGCCAGA
|| || || || || || || || || || || || || || || || || || || || || || || || || || || ||
TGAACCTGGCGAAACGGTATATGCATTGCCGTGAACCGACCAAAAGGAGGGTTCGATGAACCGC-CTACACG
1070 1080 1090 1100 1110 1120 1130

330 340 350 360 370 380 390
GCG-GCTGGCGTGCACGATTGTCGATGTGTCGACTTTGCGATATCGGTCGCGACGTCAGTTATTGGCCGA
|| || || || || || || || || || || || || || || || || || || || || || || || || || ||
GTGTCGGCGCTGGCCCTTGA-TGCCG--GGGTGAGCGTGC-ATATCGTGCACGACTAC---CTGCTGCGCG-
1140 1150 1160 1170 1180 1190 1200

400 410 420 430 440 450 460
GGTTTC-GCGGGCTGCCGACGTTCATATCGTGGCGGCGACCCGGCTTGTGGTTCGAC-CCGCCACTT---TCG
|| || || || || || || || || || || || || || || || || || || || || || || || || || ||
GATTGCTGCCGAGTCG-CCTGCACACG-GGTGGCTA-CGGCCTGTCGATGACGCCGC--CTTGCAGCG
1210 1220 1230 1240 1250 1260

470 480 490 500 510 520
A-TGCGAT---TGAGGGTATG---TAGAGGAACTCACACAGTTCTTCCTGCGTGAGATTCAATATGGCATCGA
|| || || || || || || || || || || || || || || || || || || || || || || || || ||
ACTGTGCTTCGTCGGGCCGCGCTTCAAGG--CGGGCATCGGCCT--CGGCG---CATTGGCGCGGCTGTGCC
1270 1280 1290 1300 1310 1320 1330

1260	1270	1280	1290	1300	1310	1320	
AAAGCC-GGTGCCCCACCCCTGTCGATACTCTTGAG-GGACGGTAGCGACCGACCGTGCCTTTCGTGAACGTGC							
GATGCCGGGTTGCCCGCCGCGATCTTCAAATTGAGCTGACCGAATCG-----GTTGCGTTCGGTGATCCGG	2080	2090	2100	2110	2120	2130	2140
X							
AG							
:							
CGATCTTC							
2150							

9. LOW344-FIG1. SEQ

BLYAMY2 Barley (*H. vulgare*) alpha-amylase 2 mRNA, complete

LOCUS BLYAMY2 1588 bp ss-mRNA PLN 15-JUN-1988
 DEFINITION Barley (*H. vulgare*) alpha-amylase 2 mRNA, complete cds.
 ACCESSION M17128
 KEYWORDS alpha-amylase.
 SOURCE Barley (*H. vulgare* cv Sundance), cDNA to mRNA, clone E.
 ORGANISM *Hordeum vulgare*
 Eukaryota; Plantae; Embryobionta; Magnoliophyta; Liliopsida;
 Commelinidae; Cyperales; Poaceae; *Hordeum*; *vulgare*.
 1 (bases 1 to 1588)
 REFERENCE
 AUTHORS Knox, C. A. P., Sonthayanon, B., Chandra, G. R. and Muthukrishnan, S.
 TITLE Structure and organization of two divergent alpha-amylase genes
 from barley
 JOURNAL Plant Mol. Biol. 9, 3-17 (1987)
 STANDARD full staff_entry
 COMMENT Draft entry and computer-readable sequence for [1] kindly provided
 by S. Muthukrishnan, 22-SEP-1987.
 FEATURES Location/Qualifiers
 mRNA 1.. 1588
 /note="alpha-amylase 2 mRNA (alt.)"
 mRNA 2.. 1588
 /note="alpha-amylase 2 mRNA (alt.)"
 mRNA 3.. 1588
 /note="alpha-amylase 2 mRNA (alt.)"
 mRNA 4.. 1588
 /note="alpha-amylase 2 mRNA (alt.)"
 CDS 97.. 1413
 /note="alpha-amylase 2"
 BASE COUNT 343 a 484 c 480 g 281 t
 ORIGIN 380 bp upstream of SstI site.
 Initial Score = 107 Optimized Score = 550 Significance = 10.17
 Residue Identity = 50% Matches = 714 Mismatches = 538
 Gaps = 169 Conservative Substitutions = 0

X	10	20	30	40	50		
CTGCAGCCTGACTCAGC-ACCAAGTCCTGCAAGCAGAGTCGT--AAGCA-ATCGCAAGGGGG							
GGTTGGCGTCCGGCCACC-AAGTCCTCTTCAGGGTTCACTGGGAGTCGTGGAAGCAGAGCGGGCGGGTGG	160	X 170	180	190	200	210	220
60	70	80	90	100	110	120	
--CAGCATGCAAACGAGAAGGGTTGCTCAAGTC1GCGGCCGAGAACCTC-TGCTGGCGGCCTGGCTGG							
TACAACATGATGATGGCAAGGTGACGACATCGCCGCCGTCGGAGTCACCCACGTCTGGCTGCCACCGCCG	230	240	250	260	270	280	290
130	140	150	160	170	180	190	
GTGCGCGACGTGGCTGGATCGATCAGCACAGC-GATCGGATCAATACGTGCGCGTCTATCACAAATCTCT							

10. LOW344-FIG1. SEQ

BPECYADE *Bordetella pertussis* cyaD gene 3' region and cyaE gene

LOCUS BPECYADE 2040 bp ds-DNA BCT 15-MAR-1990
 DEFINITION *Bordetella pertussis* cyaD gene 3' region and cyaE gene; proteins necessary for transport of calmodulin-sensitive adenylate cyclase-haemolysin (cyclolysin).
 ACCESSION X14199
 KEYWORDS adenylate cyclase; cya gene; cyaD gene; cyaE gene; hemolysin; secreted protein; toxin.
 SOURCE *Bordetella pertussis*.
 ORGANISM *Bordetella pertussis*
 Prokaryota; Bacteria; Gracilicutes; Scotobacteria; Aerobic rods and cocci; Alcaligenaceae; *Bordetella*; pertussis.
 REFERENCE 1 (bases 1 to 2040)
 AUTHORS Glaser,P., Sakamoto,H., Bellalou,J., Ullmann,A. and Danchin,A.
 TITLE Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of *Bordetella pertussis*
 JOURNAL EMBO J. 7, 3997-4004 (1988)
 STANDARD simple automatic
 COMMENT *source: strain 13323. see Y00545 for upstream cya gene; cya operon is organized cyaABDE. cyaB (712 aa) is initiated 78 bp downstream of cyaA stop; cyaB stop overlaps with cyaD initiation

(440 aa).

EMBL features not translated to GenBank features:

key	from	to	description
RBS	344	347	put. rRNA-binding site

FEATURES Location/Qualifiers

CDS <1..353

/note="cyaD polypeptide (AA at 3)"

CDS 355..1779

/note="cyaE polypeptide (AA 1-474)"

BASE COUNT 282 a 679 c 770 g 309 t

ORIGIN

Initial Score = 82 Optimized Score = 590 Significance = 10.17
Residue Identity = 51% Matches = 732 Mismatches = 495
Gaps = 197 Conservative Substitutions = 0

X 10 20 30 40 50 60
CTGCAGGCTG-ACTCGGCACCCAGTC---GCT-GCAAGCAGAGTCGTAAGCAATCGCAAGG-GGGC-AGCAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCGGC-GCCGGCATCCAGGTCCAGGCTCAGCTCGACAGCAAGGACATCGGCTTGTCAGGGCGGGCGCGCCA
X 10 20 30 40 50 60 70
70 80 90 100 110 120 130
GCAAACGAGAAGGGTTGTGCTCAAGTCTGCGGCCGAGGAACACTCTGCTGGGCGGCCTGGCTGGGTGCGCGAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCTACCGTCAA-BGTCG-GC---GCCTACGACTATACTGAAGTACGGAACGCTCGAAGGCAAGGTGTTGTAT
80 90 100 110 120 130
140 150 160 170 180 190 200
GTGGCT--GGAT-CGATCGGCACAGGCGATCGGATCAATACGTGCGCGT--CCTATCACAATCTCTGAAGCG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTGTCTCCGGATACGGTGGTC---GACGACCG--CCAACA-GCACTCGTACCGCGGTGACGATCGC---GCT
140 150 160 170 180 190
210 220 230 240 250 260
GGTTTACACTGAC--TCACGAGGAC---ATCTCGGCAGCTC--GGCAGGATTCTTGCCTGCTGGCCAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GG-CGCACCCCTGCCCTGGAGGTTGGACGGCAAGCCGCGGCTGCTCAAGGAAGGCAATGGCG-GTGC-AGGCC-G
200 210 220 230 240 250 260
270 280 290 300 310 320 330
AGTTCTTCGG--TAGC-CGCAAAGCTCTAGCG-GAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGGC--TG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
A--TATCCGGACCGGGCTCGCGCGCCTCATCGAGTATCTGCTCAGCCGG-TGGCGCGGCAATGCCGGCGAAA
270 280 290 300 310 320 330
340 350 360 370 380 390 400
GC--GTGCGAACGATTGTCGATGTGTCGACTTTCGATATCGGTCGCGACGTCAGTTATTGGCCGAGGTTTC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCCTGGGGAGCGCTAG--CATG-GCGG-CGGTGCAGGCGACGCGGGCGGGCCCTGGCGTTGGCGCT
340 350 360 370 380 390 400
410 420 430 440 450 460
GCGGGCTGCCGACGTTCATATCGT-GGCGCG---ACCGGCTTGTGGTTGACCCGCCACTTCG-ATGCG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTGGGCGGGTTCGCGCTGAGCGTGGGAGGCGGGGTGCGGGCGCGCGAT--GGCCTGGCAACGCCGCGCG
410 420 430 440 450 460 470
470 480 490 500 510 520 530
ATTGAGG--TATGTAGAG---GAACTCACACAG--TTCTTC-CTGCGTGAGATTCAATATGGCATCGAAGAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTCGAGGGGCCAGGCGGGCGCCTGCCATCTCGTGGCTTGTCGCCGCGCCGGCGGATC---GGC-TCGACGAC
480 490 500 510 520 530 540
540 550 560 570 580 590 600
ACCGGAAATTAGGGCG-GGCATTATCAAGTGCAG---ACCACAGGCAAGGCGACCCCCCTTCAGGAGTTAG